

**RARE EVENTS IN INTERNATIONAL RELATIONS:
MODELING HETEROGENEITY AND
INTERDEPENDENCE WITH SPARSE DATA**

by

Scott J. Cook

BA, Oakland University, 2007

MA, University of Pittsburgh, 2011

Submitted to the Graduate Faculty of the
Kenneth P. Dietrich School of Arts & Sciences in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH
KENNETH P. DIETRICH SCHOOL OF ARTS & SCIENCES

This dissertation was presented

by

Scott J. Cook

It was defended on

May 2, 2014

and approved by

Jude Hays, Associate Professor, University of Pittsburgh

Burcu Savun, Associate Professor, University of Pittsburgh

Daniela Donno Panayides, Assistant Professor, University of Pittsburgh

Robert Franzese Jr., Professor, University of Michigan

Dissertation Advisors: Jude Hays, Associate Professor, University of Pittsburgh,

Burcu Savun, Associate Professor, University of Pittsburgh

**RARE EVENTS IN INTERNATIONAL RELATIONS:
MODELING HETEROGENEITY AND INTERDEPENDENCE WITH
SPARSE DATA**

Scott J. Cook, PhD

University of Pittsburgh, 2014

The interdependence of international events is obvious to even casual observers of global politics. History is replete with examples of events repeating within states and/or being influenced by outcomes in other states. Despite this, much of the current literature in International Relations either mishandles or outright neglects this dependence, thereby threatening the credibility of our inferences. In large part, this stems from the difficulty of modeling such dependence when one's data are binary and rare, as they often are for many of the most widely-studied phenomena in IR (e.g., violent conflict, economic crises, etc. . .). For data of this type, commonly-used strategies to capture dependence are frequently ill-suited and, as such, new approaches are required. Therefore, this thesis aims to clarify the empirical challenges which arise from these data, detail the problems with existing approaches, and offer alternatives which should be preferred.

The focus is principally on two potential (and related) sources of bias which may arise within binary time-series cross-sectional (b-TSCS) data: true (inter-)dependence and unit heterogeneity. In the first, the outcomes, actions, and/or choices of some unit-times depend directly on those of other unit-times. To model both spatial and serial dependence in such data, a spatiotemporal-lag probit model estimated using maximum-simulated-likelihood using recursive-importance-sampling (MSL-by-RIS) is presented. This allows us to directly model the dependence of the lagged-latent outcomes, which is shown to have several advan-

tages over models using the observed indicator (e.g., model consistency, effects estimation, predictive accuracy). The second main focus is on the threat of unobserved unit heterogeneity, that is, when time-invariant unit-characteristics influence the outcome, action, choice, but go unmodeled. While fixed-effects estimators are traditionally the solution to this issue, such models have received heavy criticism in political science applications with b-TSCS data. In light of these criticisms, a penalized-maximum-likelihood fixed effects (PML-FE) model is proposed which suffers from few of these drawbacks and permits the estimation of novel unit-specific substantive effects. In addition, original analyses into intrastate conflict and financial crises are offered to highlight the value of these approaches for testing existing, and motivating new, theories of international behavior.

TABLE OF CONTENTS

PREFACE	ix
1.0 INTRODUCTION	1
2.0 (INTER)DEPENDENCE ACROSS TIME AND SPACE	19
2.1 Spatiotemporal Dependence	21
2.2 MSL-by-RIS: Estimation	27
2.3 MSL-by-RIS: Effects	30
2.4 Discussion	35
3.0 THE DYNAMICS OF CIVIL WAR	38
3.1 A Regional Theory of Conflict Traps	41
3.2 Research Design	47
3.3 Results	50
3.4 Discussion	58
4.0 MODEL MISSPECIFICATION: POLITICS AND THE CONTAGION OF FINANCIAL CRISES	61
4.1 Common Political Fundamentals and Signal-Extraction Failures	64
4.2 Research Design	72
4.3 Results	76
4.4 Discussion	85
5.0 UNOBSERVED HETEROGENEITY AND RARE EVENTS	87
5.1 Penalized Maximum Likelihood	89
5.2 Results	95
5.3 Discussion	106

6.0 THE KNOWN UNKNOWNNS OF CIVIL WAR	110
6.1 Cum Hoc Ergo Propter Hoc: Low Income and Civil War?	113
6.2 Rare Events and Rarely Changing Regressors	118
6.3 Developmental Peace	130
6.4 Discussion	134
7.0 SUMMARY AND CONCLUSION	136
8.0 BIBLIOGRAPHY	140

LIST OF TABLES

2.1	Simulation Results for MSL-by-RIS Coeff. Est.	30
3.1	Dependence of Civil Conflict in Sub-Saharan Africa	51
4.1	Bank Crisis Contagion (Spatial Probit), 1970 – 2007	77
4.2	Variation in Estimation Approaches	80
5.1	Coef. Est. Pooled Models ($N=50, T=20, \alpha_i \sim N$, 1000 trials)	98
5.2	Coef. Est Panel Models ($N=50, T=20, \alpha_i \sim N$, 1000 trials)	99
5.3	Average Unit Effect Estimates	106
6.1	Panel Estimators with Rarely Changing Regressors ($\alpha_i \sim Normal$)	124
6.2	Panel Estimators with Rarely Changing Regressors ($\alpha_i \sim Bimodal$)	126
6.3	Type 1 & Type 2 Errors with Slowly-Changing Regressors	128
6.4	Unit Effects and Civil War	131
6.5	Unit Effects and Civil War: Sample Size	133

LIST OF FIGURES

2.1	Accuracy of RIS for Long-Run Response Paths	32
2.2	Estimating Spatial Effects via Simulation	34
3.1	Global Incidence of Civil Conflict, 1950-2000	42
3.2	The Dependence of Conflict in Space and Time	46
3.3	Contagion: Contemp. Spatial Effects of Conflict(Burundi & Rwanda)	55
3.4	Contagion: Regional Spatial Effects of Conflict(Rwanda, Burundi, & DRC)	56
3.5	Conflict Trap: Persistence & Contagion (Rwanda, Burundi, & DRC)	57
4.1	Financial Crises, 1990-1994 (Mexican Peso Crisis and ERM)	66
4.2	Financial Crises, 1995-1999 (Asian Financial Crises)	66
4.3	Competing Theories of Contagion (Spillovers vs. Informational)	67
4.4	Spatial Relations Between Mexico and the World	81
4.5	Scatter Plot of Reduced Form Disturbances (Guatemala & Mexico 1994))	82
4.6	Short and Long Run Effect of Shock' (Guatemala and Mexico, 1994 2000)	83
4.7	Short and Long Run Effect of 'Shock' (US and Mexico, 1994 1997)	84
5.1	Accuracy of Coefficient Estimates ($\alpha_i \sim N$)	101
5.2	Accuracy of Coefficient Estimates ($\alpha_i \sim \chi^2$)	103
5.3	Accuracy of Coefficient Estimates ($\alpha_i \sim Bimodal$)	103
5.4	Accuracy of MEMs ($\alpha_i \sim N$)	105
5.5	Accuracy of MEMs ($\alpha_i \sim \chi^2$)	105
5.6	Accuracy of MEMs ($\alpha_i \sim Bimodal$)	105
6.1	Development and Civil War Incidence, 1950-2008	114

PREFACE

Though most will surely never read this, there are many people who were essential in helping me to reach this point. Of course, my family and friends, for the many times they have been there for me throughout life. Mary Wilson, for her love and understanding, particularly as I was in the final stages of writing this. The many teachers who encouraged or inspired me to choose this path. Notably, Jeff Lopo, for first sparking my interest in politics, and Peter Trumbore, for making me believe that I could make a future out of it.

While at Pitt, I was extremely fortunate to be surrounded by many helpful and encouraging people, both fellow struggling graduate students and understanding faculty. I owe a particular debt to Professors Burcu Savun and Jude Hays. Thank you, Burcu, for mentoring me through all that it takes to be an academic. From my first conference to my first practice job talk, you have always been there for me and I am so thankful for that. Then, as if I wasn't lucky enough to have one great advisor, I got another in my third year who completely changed my perspective on political science, and inspired me to pursue a completely different type of research. Jude, thank you for your wisdom and your patience, I am truly grateful to have had the opportunity to work with, and learn from, you. This project simply wouldn't have been possible without your guidance.

Lastly, to my mom, Sandra Cook, an amazing woman to whom I owe everything. Above all, this dissertation is dedicated to you.

1.0 INTRODUCTION

I don't particularly care about the usual...Can you assess the danger a criminal poses by examining only what he does on an ordinary day? Can we understand health without considering wild diseases and epidemics? Indeed the normal is often irrelevant.

— Nassim Taleb

...the opportunity to be wrong is considerably enhanced when the design is two-dimensional.

— James A. Stimson

Many of the most interesting phenomena in International Relations, and political science more generally, are either by nature or design binary. Which is to say that the choices, actions, and outcomes of states can only be observed as having occurred or not, as being present or absent. Binary outcomes inherently complicate our ability to understand what gives rise to these phenomena, as they reveal less information about the process by which they were produced than do continuous outcomes.¹ By analogy, it is easier to estimate the utility a state places on military preparedness (e.g., military consumption over GDP), than it is to estimate the utility it places on avoiding war (i.e., war/no war). While this concern is common to any analysis with binary outcomes, it is more salient in International Relations where our binary outcomes are almost invariably *rare* as well. When events occur infrequently, we have even less information from which to derive our understanding of their causes. Despite these unique challenges, surprisingly little attention has been paid to

¹[Train \(2009\)](#) discusses this in terms of deriving the choice probabilities for individuals when making a selection.

the consequences of estimating models with binary rare-event data. In part, this is because other fields do not as often confront these data in their analyses.² However, rare and binary outcomes are *central* in International Relations, with models of such events representing amongst its most significant and substantial contributions.

Perhaps nowhere is this more apparent than in the study of the causes of interstate war. Despite the (thankful) rarity of such wars, little work in political science has had the impact of the democratic-peace theory (Doyle 1986, Maoz & Russett 1993, Russett 1994).³ That democracies rarely go to war with each other is taken as one of the law-like empirical regularities of political science (Levy 1988). In its wake, a substantial amount of research has sought to better understand (e.g., De Mesquita *et al.* 1999) or explain away this finding, arguing that it is a consequence of other related factors (e.g., alliances (Gowa 1994; 1995), capitalism (Gartzke 2007), American hegemony (Rosato 2003)) or model misspecification (Green *et al.* 2001), though evidence continues to offer support for the peace between democracies (Dafoe *et al.* 2013). In addition to motivating this line of research, the democratic-peace theory is unique among political-science research in that it seems to have carried meaningful weight in foreign policy decision-making. Presidents Clinton and Bush both expressed a belief that the spread of democracy was an essential component for the promotion of peace, under the logic that democracies don't attack one another. As a result, it has been used as a motivation for the promotion of democracy and, it is argued, as a potential justification for war itself (e.g., 2003 Iraq War).

Other forms of conflict are also quite rare, including the study of intrastate war. While civil war is now the most frequent and prominent form of armed conflict, it remains a rare event. As Gates (2002) notes an “inherent problem characterizing civil war data is the relative rareness of this event,” (22). This is particularly true if we confine our attention to new wars, of which there are rarely any in a given year. However, when we do see civil war

²Of course there are notable exceptions to this. In biostatistics and medicine low rates of occurrence in randomized trials often requiring cross-trial clustering procedures, in behavioral economics the importance of rarity in shaping expectations (and responses) has been widely discussed, in machine learning and language processing algorithms for divining patterns from extremely sparse data are increasingly common. The more fundamental point is that these events are not so central in other fields so as to have motivated the construction of a range of general techniques and strategies when dealing with models of rare events.

³Beck (2008) notes that Maoz & Russett (1993) is the second most widely cited article from the American Political Science Review in the last twenty five years.

it is often in states within the context of recurring or enduring conflicts. Consequently, a small fraction of states account for the majority of civil wars that are observed, motivating researchers to attempt to understand why conflict is so frequent in these countries but so rare everywhere else (Collier *et al.* 2003). The dominant explanation given in the literature is that underdevelopment in these states gives rise to conflict, creating conditions conducive to rebellion (Collier & Hoeffler 2004, Fearon & Laitin 2003). As with the democratic peace for interstate war, the ‘developmental peace’ is one of the most robust empirical regularities in the discipline, with some arguing that it has resolved the debate on the causes of civil war (Rice *et al.* 2006). As a result, scholars consistently advocate and policy-makers have increasingly begun to pursue development as a strategy for reducing the risk of civil war. Given the severe consequences of civil war – casualties in the millions, destabilizing refugee outflows, domestic and regional economic costs in the millions, disease, pollution, etc... – it is clearly important to understand why they occur and, in turn, what might be done to make them even rarer.

Rare-event outcomes are not unique to peace studies, however, research into the emergence of economic regionalism is among the more important work in international cooperation. According to the World Trade Organization (WTO), 377 regional trade agreements (RTAs) were in force as of 2014. While this is less rare than interstate wars, as a proportion of the total number of dyads which could have agreements it is still quite rare. Researchers have sought to understand the domestic and international political conditions affecting the decision to enter into RTAs, arguing for the impact of waning American hegemony (Mansfield & Milner 1999), stalled multilateralism (Mansfield & Reinhardt 2003), joint democracy (Mansfield *et al.* 2002), and domestic veto-players (Mansfield *et al.* 2007). Given the purported benefits or costs of these agreements suggested by the literature for trade (Magee 2008), investment (Büthe & Milner 2008), repression (Hafner-Burton 2005), and war (Mansfield & Pevehouse 2000), it is important to understand the conditions which cause states to enter them with the states they do.

International Political Economy scholars also frequently analyze rare events, with the study of economic crises being the most recent and prominent example. As noted by (Leblang

& Satyanath 2006, 245) “currency crises are are costly phenomena that have been exceptionally difficult to explain and predict.” These rare events result in severe economic decline (Bordo *et al.* 2001), often crippling economies for years after, and social and political unrest (Frankel 2005), including shorter leader tenure and regime turnover. Banking crises are equally debilitating, imposing considerable fiscal costs (Laeven & Valencia 2012). Furthermore, banking crises often precipitate currency and/or sovereign debt crises in their wake (Reinhart & Rogoff 2011). While our understanding of the causes of the events remains limited, IPE scholars have increasingly offered political explanations for these events. Notably, scholars have argued that political institutions (e.g., veto players, divided government, government turnover) determine both the ability of states to respond to crises and the expectations of investors about these responses Leblang & Satyanath (2006), MacIntyre (2001). Given the recent global recession it is likely that analysis into these events will only increase in the future Helleiner (2011), Mosley & Singer (2009).

As can be seen, rare and binary events are routinely found in the study of International Relations. Therefore, it is important to know what, if any, unique challenges are raised in their analysis. We have already mentioned that they provide less information, but how exactly does that matter? Returning to the democratic-peace literature helps illustrate the possible problem(s). Critics have suggested that one of the impediments to gaining an accurate understanding of the relationship between democracy and conflict is the rarity of war itself.⁴ Spiro (1994) was among the first to voice these concerns, arguing that the failure to observe democracies at war is not theoretically meaningful, but instead a function of the rarity of war and the paucity of jointly democratic dyads. In particular, he expressed, if somewhat imprecisely, two concerns about Maoz & Russett (1993)’s analysis which continue to cloud the democratic peace debate today: *i*) estimating a model of war which pooled all dyad-years and *ii*) treating each dyad-year as if it were independent. While much progress has been made on these issues since, these concerns remain fundamentally the same problems faced by researchers today.⁵ In general, as the number of occurrences of an event decreases, it

⁴This is in addition to a range of other measurement issues – primarily centered on ones understanding and coding of democracy – which will not be addressed here.

⁵These advances include the efforts to better account for temporal dependence in binary data (Beck *et al.* 1998, Carter & Signorino 2010), and a thorough discussion on the appropriateness of various panel models

becomes increasingly difficult to determine whether our empirical findings are a consequence of substantive theoretical explanations, (un-modeled) heterogeneity among the units, or dependence within the outcome. Given these constraints, researchers are often forced to choose between several second-best choices, often with little guidance or justification. As a consequence, even those relationships widely regarded as “empirical law” rest on potentially specious modeling assumptions.

Consider again the work on civil war onset, with its focus on explaining the uneven distribution of conflict among states. It could be, as is argued by the developmental peace literature, that the states which experience conflict regularly are simply abundant in the factors that produce conflict (e.g., low GDP per capita).⁶ However, there are alternative explanations which would place these conflicts in the same countries. First, these states are also likely to possess *unobservable* factors which make civil war more likely (e.g., weak social institutions, historical antipathy among ethnic groups, etc...). Moreover, these same unobservable factors may also be a cause of the low levels of development in these countries, complicating efforts to understand the direct impact of development on civil war (Djankov & Reynal-Querol 2010). Second, given that these countries experience repeated episodes of conflict it would seem to indicate dependence between these events. This is exactly the point raised by the ‘Conflict Trap’ literature, which suggests that experiencing a civil war makes subsequent civil wars more likely due to economic destruction, heightened antipathy, and the introduction of guns, troops, and ideologies of conflict into the state (Collier *et al.* 2003). As such, it is argued that the dynamics of conflict itself produces additional conflict. Furthermore, civil war also generates lower levels of development, which again calls into question the meaningfulness of the observed relationship between GDP and conflict. While conflict scholars appreciate these alternative explanations, their ability to discriminate between them has been constrained by limitations in the available techniques for modeling binary and rare event data.

(Beck & Katz 2001, Green *et al.* 2001, King 2001, Oneal & Russett 2001). Both advances will be discussed in greater detail in subsequent chapters.

⁶Here and throughout the remaining chapters I use the ‘developmental peace’ to succinctly refer to the various works linking low GDP to civil war. I simply note this as it is not a phrase which is common to the literature (or at least not that I am aware).

In sum, the rate at which we have produced theories pertaining to rare phenomena has seemed to outpace our ability to accurately estimate models of rare data. In the broadest sense, binary data confront many of the same threats to accurate inference found with interval data, but researchers possess fewer solutions for dealing with and overcoming these concerns. Moreover, the rareness of these data itself complicates the efficacy of even existing strategies explicitly suited for binary outcomes. These methodological shortcomings and my belief in the importance of the substantive areas affected by them serve as the motivation for the remaining chapters. In short, I will discuss the nature of some of the empirical challenges presented in models of binary data, provide possible solutions to these issues, and indicate how these strategies enable us not only to test our existing theories more accurately, but actually motivate (and demand) new theories of international behavior. Moreover, I present original analyses into the study of war and financial crises which challenge conventional understandings of these events. To help frame these issues, I open here with a brief discussion of the conventional design and estimation of binary-outcome models. This also allows me to introduce some of the notation which will be used throughout.

While most of the empirical work on binary and rare events in International Relations increasingly utilizes time-series-cross-sectional (TSCS) data – which is defined loosely by [Beck \(2008\)](#) as “a relatively small number of units observed for some reasonable length” – some questions remain ill-suited for, or lack the data demanded by, TSCS analysis.^{7,8} Therefore, for both generality and ease of exposition, I begin the discussion focusing exclusively on cross-sectional analysis before expanding the analysis to include a time dimension.⁹ A simple model for cross-sectional data with a binary outcome is presented in [Equation 1.1](#). While there are several analogous ways to motivate the generating process for binary data,

⁷[Beck \(2008\)](#) partly credits the success of the “Democratic Peace” literature for the proliferation of time series cross-sectional data.

⁸For example, the use of child soldiers in civil war is predominantly modeled with “conflict” as the unit of analysis.

⁹I do not discuss time-series analysis uniquely as it is rare for empirical IR to exhibit no cross-sectional variation. An exception would be theories on changes at the systemic level which examine only variation across time. For these contexts the later discussion on modeling temporal dependence in TSCS data should suffice.

the latent-variable representation is particularly intuitive and will be used in the remaining chapters.^{10,11} Under this formulation the regression function is given by:

$$y_i^* = \alpha_i + \mathbf{X}_i\boldsymbol{\beta} + \epsilon_i \quad (1.1)$$

where α is the constant, \mathbf{X}_i is k matrix of included regressors, $\boldsymbol{\beta}$ is a vector of their associated coefficient-parameter estimates, ϵ_i is an i.i.d error disturbance, and identifying subscript $i = \{1, \dots, N\}$ denotes the unit (observation). Latent-variable y^* links to observed y through the measurement equation:

$$y_i = \begin{cases} 1, & \text{if } y_i^* > 0 \\ 0, & \text{if } y_i^* \leq 0 \end{cases} \quad (1.2)$$

If instead latent-variable y^* were observed, estimation would be straight-forward, as it is a linear function of the regressors, with ϵ_i giving the prediction errors. Instead, we only observe y^* in terms of its sign (1.1) and therefore must specify a link function to relate the distribution of the observed responses to the linear predictors. The two common link functions used to estimate binary outcomes are logit and probit.^{12,13} Both will provide the response probability (conditional on \mathbf{X}_i that we are interested in:

¹⁰The latent variable formulation will also be useful later in discussing the correlations among observations.

¹¹Alternatives include the index model, the utility model, etc...

¹²In applied empirical International Relations work logit is clearly the preferred choice. While some may be making this choice purposefully – the fatter tails of the logistic CDF are argued to make it less susceptible to outliers which may be more present in rare-event data – in practice there is often little justification provided.

¹³Another approach to binary data which is widely discussed in the econometric literature but is not commonly used in International Relations is the linear probability model [Angrist & Pischke \(2008\)](#). While the LPM is easy to estimate – simply regressing y on \mathbf{X} using OLS – the predicted errors are necessarily heteroskedastic and frequently fall outside the unit interval, therefore such approaches are typically cautioned against in most introductory econometrics texts ([Wooldridge 2012](#)). More seriously, as Dave Giles has noted, LPM cannot produce consistent estimates of the true marginal effects as it does not produce consistent estimates of the parameters. Furthermore, any measurement error – mis-classification of zeroes and ones – has been shown to be a much more serious problem for LPM models than logit or probit ([Hausman et al. 1998](#)). Mis-classification is not limited to an incorrect coding of the data, but rather is also present in situations where units are modeled as being at risk of failure when they are not truly. As researchers have argued in a different context – as motivation for split-population models ([Xiang 2010](#)) – this is likely the case for a number of the more prominent applications in international relations (e.g., interstate war initiation). Given these limitations and the ease with which one can estimate probit and logit models using conventional statistical packages, I do not discuss the use of the LPM model of binary data any further.

$$Pr(\epsilon_i < \mathbf{X}_i\boldsymbol{\beta}) = Pr(y_i = 1|\mathbf{X}_i\boldsymbol{\beta}) = \pi_i \equiv F(\mathbf{X}_i'\boldsymbol{\beta}) \quad (1.3)$$

With differences arising from the specification of Cumulative Distribution Function (i.e., $F(\cdot)$), with probit specifying the errors as distributed standard-normally and logit distributed logistically:

$$\begin{aligned} \textbf{Logit:} \quad \pi_i &\equiv \Lambda(\mathbf{X}_i\boldsymbol{\beta}) \equiv \frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta})} \\ \textbf{Probit:} \quad \pi_i &\equiv \Phi(\mathbf{X}_i\boldsymbol{\beta}) \equiv \int_{-\infty}^{\mathbf{X}_i\boldsymbol{\beta}} \phi(v)dv \end{aligned} \quad (1.4)$$

With the joint density of y_i for all i given \mathbf{X}_i for each logit and probit being the product of the individual observation's density:

$$f(y|\mathbf{X}_i\boldsymbol{\beta}) = \prod F_i^{y_i} (1 - F_i)^{1-y_i} \quad (1.5)$$

and the log-likelihood the sum of the individual log-likelihoods:

$$\log \mathcal{L}(\boldsymbol{\beta}) = \sum \{y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i)\} \quad (1.6)$$

which is then maximized to provide parameter estimates $\hat{\boldsymbol{\beta}}$ which best fit the data. Assuming the observations are independent and normally distributed $\hat{\boldsymbol{\beta}}$ is a consistent, asymptotically normal, and efficient estimate of $\boldsymbol{\beta}$, and is produced via the standard pooled logit or probit estimator. Unfortunately, these assumptions are rarely valid even in cross-sections of data, as our observations (and therefore marginal probabilities) are often correlated *across* units.

More informally, such (inter)dependence may be present whenever we have cause to suspect that the actions, outcomes, and choices of states (dyads, etc...) are influenced by the choices of some other actors. Which is to say, *always and everywhere*. Indeed coming up with

examples where the actions of states are completely independent is far more difficult than thinking of situations where they are affected by one another. Even more to the point, what is the field of International Relations if not a collection of theories on how the interactions between states produce outcomes which would otherwise be different? Historically (and colloquially), interdependence has been used to describe those situations in which members opt-in to some system which results in a form of intended mutual dependence (e.g., trade, alliances, agreements). These connections, however, also create a structure through which the actions taken in one member-state i have consequences for another member-state j . More accurately, given the ubiquity and density of these ties in the current global landscape, the actions of i affect *all* other states $j \neq i$, either directly or indirectly. Furthermore, and somewhat obviously, the likelihood with which some event influences others increases as a function of the significance of the event.¹⁴ This means that the very issues which are presumably the most interesting to scholars (i.e., the important ones) are also those which are the most likely to be affected by spatial interdependence.¹⁵

As [Franzese & Hays \(2007; 2008\)](#) have shown analytically and numerically, failing to account for this spatial dependence results in inefficient and inaccurate standard errors at best, and generally bias as well. To avoid this, we will typically want to include a spatial-lag – analogous to the more familiar temporal lag except that it relates the outcomes of different units weighted by some measure of relation (canonically distance) between the two (Chapter 2 provides a more formal treatment) – to capture the interdependence among the units as follows:

¹⁴Drawing upon the Epidemiologic Triad for disease transmission, we might consider the risk of spatial spillovers to be a function of three factors: the virulence of an external agent, the susceptibility of host, and the conditions of the environment, each of which interact to determine whether a disease develops. There are obvious parallels to this in our understanding of the spread of political phenomena and it is a model I develop more fully later.

¹⁵I should note that the likely presence of and theoretical consequences which result from such interdependence are increasingly being explored by International Relations scholars in most all issue areas. [Simmons et al. \(2006\)](#)’s work identifying four distinct theoretical mechanisms for the diffusion of liberalism – *i*) coercion, *ii*) competition, *iii*) learning, and *iv*) emulation – has been widely cited and spurred work on diffusion in a range of additional issue areas. Furthermore, studies exploring the spread of social movements, the diffusion of regimes, and the contagion of conflict are increasingly common.

$$y_i^* = \alpha_i + \rho \mathbf{W}\mathbf{y}^* + \mathbf{X}_i\boldsymbol{\beta} + \epsilon_i \quad (1.7)$$

With continuous outcomes, we then simply estimate this model using one of two consistent and relatively easy to implement estimators (Spatial-ML and Spatial-2SLS).¹⁶ However, as briefly introduced in [Franzese & Hays \(2008\)](#) and elaborated (& extended) in [Franzese *et al.* \(2014\)](#), estimating spatial models with qualitative dependent variables is not as straightforward. In short, spatial interdependence renders standard probit inappropriate as the marginal probabilities are no longer independent, meaning that the joint probability is not simply the sum of the log of marginal probabilities (as in [Equation 1.6](#)).¹⁷ As such, estimators which either estimate non-spatial or ‘naïve’ spatial models – the only models currently employed for models of binary and rare events in International Relations – will produce potentially wildly biased estimates.

The likelihood of dependence in our data only increase if we introduce over-time data to our cross-section. While the myriad benefits that come from time-series-cross-sectional data have rightly made it the dominant “large- N ” data structure in International Relations, it becomes even less likely that our observations are independent. That is, to estimate a conventional logit or probit – one with no correction for (latent) auto-correlated residuals – we would have to be willing to assume that the likelihood with which a state experiences an event in one period is unrelated to the likelihood that a states experiences that event in other periods, which is rarely, if ever, the case. To see these issues, I now expand our model from [Equation 1.1](#) to include variation in the outcome across time:

$$y_{it}^* = \alpha_i + \mathbf{X}_{it}\boldsymbol{\beta} + \epsilon_{it} \quad (1.8)$$

with t indexing time $t = \{1, \dots, T\}$ and it now indexing (unit-time) observations $it = \{1, \dots, NT\}$. As with spatial auto-dependence, any dependence within a single unit i across

¹⁶[Franzese & Hays \(2007\)](#) even indicate that Spatial-OLS performs reasonably with low levels of interdependence.

¹⁷As discussed later, we must instead maximize the log of one n-dimensional non-separable probability.

time t will complicate estimation. Namely, it would mean that, once again, our marginal probabilities are not independent. While this is unlikely to result in bias, as in the case of spatial interdependence, serial dependence in the data still results in inefficient and incorrect standard errors, thereby risking overconfident inferences ([Beck 2008](#), [Poirier & Ruud 1988](#)). This dependence manifests frequently in many of the rare events we have discussed thus far, with many exhibiting substantial persistence in the outcome. As an example, [Gates \(2002\)](#) notes that ‘conflict data tend to be characterized by complex dependence structures...any econometric analysis of civil conflict must account for this lack of independence across cases,’ (22). The same is true for almost anything of interest.

It is, therefore, important to account for both a state’s previous actions and its ‘neighbors’ current ones, as these outcomes condition the future behavior of a state.¹⁸ These ‘prior’ events (in time or space) alter the factors which determine future occurrences. We see evidence for the logic of this framing every time we make a decision: my choice today is a consequence of my prior experiences. However, another explanation could be also be linking these decisions, perhaps I am just the type of person who is inclined to make a particular choice (regardless of previous experiences). As noted by [Heckman \(1978\)](#), the prior explanation assumes an implicit counter-factual wherein the same actor would have a different probability of experiencing an event (making a choice, etc...) today if the past had been different, that is, if that event, choice, action had not occurred. This is true dependence. Another explanation is that we see persistence in the actions taken by individual actors because they have heterogeneous propensities to experience the event which are themselves correlated across time. In this respect, the prior and current events do not meaningfully depend upon one another, but are only related to the extent that it is the same actors experiencing them. As any treatment on panel data informs us, this unobserved unit-level heterogeneity may pose the same threat to inference on the substantive parameters as dependence did in our prior discussion. As such, it is important to account for it in our analysis.

In short, we need to establish baseline expectations for the behavior of all units in the sample (i.e., constants). Different assumptions about the heterogeneity of the time-

¹⁸Neighbors used literally here, but only as a convenient term to denote spatial proximity.

invariant latent propensities (and ultimately their relation to the included regressors) give rise to different models. If one assumes that $\alpha_i = \alpha$ – that is, no unit level heterogeneity – one simply estimates pooled probit or logit models. To the extent that the $\alpha_i = \alpha$ these models are unbiased and provide the most efficient estimates of the parameters. But notice how restrictive this assumption is, it requires that there are no unobserved factors which distinguish between the units in the model. This is rarely the case. Furthermore, when significant differences between α_i and α exists – i.e., when there is unit-level heterogeneity – this model is misspecified and the estimates will be biased. Often it will be more realistic to assume that α_i varies across units and specify a model which accounts for this unobserved unit-specific variation.¹⁹ This gives rise to two broad estimation strategies – distinguished by the assumption about the relationship between α_i and \mathbf{X} – random and fixed effects. In short and saving the technical details for later discussion, if one believes that the individual level effects α_i are unrelated to \mathbf{X} , we can regard them as a problem only of the stochastic term and estimate the random-effects model. Generally, however, it is unlikely the unobserved factors which affect the outcome will be orthogonal to the observed factors. If they are correlated then we need to account for the unobserved effects in the structural term and estimate a fixed effects model.

The permissiveness of these approaches for rare-event binary-outcome models in International Relations has been subject to much debate (Beck & Katz 2001, Green *et al.* 2001, King 2001, Oneal & Russett 2001). While all agree that there are likely unobserved unit effects which correlate with the regressors, many still feel fixed effects is almost never a good idea with these data. First, given that it eliminates between variation from the analysis, fixed effects models cannot provide estimates for time-invariant regressors and may provide incorrect estimates for those that are nearly time-invariant. This latter point is particularly important, as many of the explanators International Relations scholars are interested in change ‘slowly’ – that is, exhibit greater variation between units on average than within units over time – such as institutions, population, development. The other major problem with available fixed-effects models of binary outcomes is that those units which do not

¹⁹I pause to remind readers that this was the second major issue raised by Spiro (1994) regarding the empirical work on the Democratic Peace, that is, whether or not it is valid to estimate pooled models which treat all dyad-years as having a common baseline propensity for war.

change state – e.g. $\sum_t y_{it} = 0$ or T – contribute nothing to the likelihood (returning estimates of $\pm\infty$) and are consequently dropped from the analysis. As such, researchers have argued that estimating these models induces sample-selection bias by removing all the cases which never experience the outcome. Given that there are many such units in rare event data this problem is concerning. Thus, researchers with rare-event binary-outcome data face a Morton’s fork: either assume the presence of no or orthogonal individual unit effects in situations where such assumptions are likely invalid *or* estimate models which account for this heterogeneity but induce other forms of bias into the analysis (e.g., sample selection, incidental parameters, rarely-changing regressors).

In sum, I have presented two ways in which unit heterogeneity impedes our ability to make sound inferences with rare events BTSCS data. First, units have distinct histories upon which the propensity for a current event depends. This includes both prior realizations of the event within that unit itself and the experiences of other units with which that state is uniquely related. As such, our individual observations are not independent, but instead conditional upon one another. Second, units have distinct unobservable characteristics which determine their propensity to take an action. That is, even after accounting for all the theoretically relevant observed factors, units still have different probabilities of experiencing the event. As a result, events will occur in some units consistently more (less) than would be expected by the observed factors alone. Both of these issues impair our ability to draw credible inferences about the relationship between the substantive parameters and the outcome. Yet, strategies to address these issues for models of rare events binary outcomes remain limited, threatening the validity of our research into topics such as interstate conflict, civil war, and economic crises. While we will likely never find a perfect solution for handling rare-event BTSCS data, the importance and centrality of these phenomena to International Relations should compel us to improve upon those current strategies which are clearly imperfect.

Therefore, I take up the issue of dependence in rare event binary data in [chapter 2](#). Drawing on [Franzese *et al.* \(2014\)](#), I discuss the econometric challenges of these models more formally, introduce a simulation-based strategy for their estimation, and provide a novel simulation-based strategy for estimating conditional counter-factual substantive effects. This

approach greatly improves on standard approaches for estimating spatial models employed within the IR literature. Furthermore, I discuss how the same estimator can be used to model temporal dependence as well. While other strategies exist to remove the nuisance of temporal dependence, no current approach directly capture the auto-dependence in the outcome. Doing so gives a natural interpretation to the effect of the substantive parameters (e.g., short-run, long-run), in a manner that is familiar to researchers from dynamic interval-level models.²⁰ Furthermore, it offers the first fully integrated and consistent estimator of the *total* dependence for binary outcomes presented in political science. I conclude this chapter with a discussion on the applicability of these, and the more standard methods, for rare events explicitly. Additionally, I note the additional complications raised when extending this model to account for conditional and interdependent *responses* (in addition to conditional outcomes) as well. Given that we will frequently want to model these kinds of dynamics as well, I suggest it as fruitful area for future research.

In [chapter 3](#), I bring these strategies to bear on the question of civil war incidence. As noted above, civil wars tend to concentrate heavily in a particular set of states. Considerable research has sought to explain this empirical reality, with many scholars arguing that such clustering arises from auto-dependence in conflict itself. In particular, there are two ways in which civil war is likely to cause additional conflict: persistence and contagion. The literature on conflict persistence argues that fighting creates conditions favorable to additional conflicts in *that* state, whereas the literature on conflict contagion argues that civil war increases the risk of additional conflicts in *neighboring* states. Reviewing these literatures, I argue that they are part of the same systemic process of conflict dynamics and, therefore, need to be jointly analyzed. Specifically, I argue that contagion is a mechanism that increases the persistence of conflict, as each state’s risk of conflict is augmented by those of their neighbors, and more persistent conflicts, in turn, increase the likelihood of contagion. These mutually-reinforcing positive-feedback loops help explain why we observe conflict-prone regions and

²⁰Throughout I use ‘substantive’ to refer to the included regressors (assumed to be) exogenously introduced into the equation of y^* . Elsewhere, [Franzese & Hays \(2007\)](#) have referred to these as “domestic, exogenous-external, or context-conditional” effects. The use of the term substantive is primarily used in relation to the ‘nuisance’ parameters introduced in later chapters for fixed-effects estimation, I retain the term throughout for consistency, but note that it is not meant to suggest anything about the substantive *importance*, or lack thereof, of the dependence parameters.

offer a more complete understanding of the ‘conflict trap’ which ensnares states. Using the method developed in [chapter 2](#) I test for the presence of these conflict dynamics in Sub-Saharan Africa. The results suggest that there does appear to be regional conflict dynamics which arise from both the persistence *and* contagion of conflict.

While civil wars spread almost exclusively to geographically proximate countries, states can be ‘connected’ in a variety of ways. These relationships are inherently complex, as states find themselves bound together by a series of overlapping, direct (e.g., alliances, trade, etc...) and indirect ties (e.g., shared history, common language, similar polity, etc...). As such, [chapter 4](#) briefly discusses how models which employ a single measure of spatial proximity – classic SAR and STAR models – in International Relations will often be under-specified, thereby underestimating the strength of spatial interdependence. Instead, researchers should estimate multiparameteric spatial models (as previously discussed by [Hays *et al.* \(2010\)](#)), which allow for the inclusion of several measures of spatial (i.e. cross-unit) proximity and, therefore, better capture the interdependence in the data. Furthermore, it enables us to discriminate between possible sources of interdependence and, therefore, gain greater leverage over competing theoretical mechanisms. Utilizing this approach, I analyze how even *perceived* ties between states can cause their fates to be wed during times of crisis.

Specifically, I argue that contemporary explanations for the spread of financial crises – which primarily focus on trade and financial ties – have underestimated the role that investor beliefs, and subsequent behavior, play in spreading crises above and beyond what would be anticipated from direct ties. I contend that following a crisis, investors update their beliefs as to the (possibly spurious) causes of the crisis and subsequently withdraw investment from other states they believe to be at similar risk (i.e., signal-extraction failure). Building on theories of investor ‘lumping,’ I argue that states with common political fundamentals suffer from contagion even in the absence of direct economic ties. Specifically, the initial crisis induces updating over the ability of its political institutions to prevent a crisis, precipitating withdrawals from states with similar political environments and resulting in additional crises. Ultimately, I find evidence that states with similar political institutions are more likely to

experience simultaneous financial crises, suggesting that IPE scholars may need to expand their understanding of how these institutions matter.

In [chapter 5](#), I turn to the issue of unit-level heterogeneity and the challenges it presents for models with binary-outcome data in International Relations. Many of these issues emerge principally *because* our data are rare as well. Therefore, I begin with a more general discussion of what we mean by rare events and how rarity might potentially impact our analyses. [King & Zeng \(2001a;b\)](#), amongst others, have noted that because rare events provide so little information our estimates are frequently biased. While this is true for any model, the impact of rare events becomes even more pernicious when we attempt to estimate traditional panel models. With rare events it is increasingly likely that some units do not experience the outcome, meaning they are dropped from fixed-effects models and our results are potentially biased ([Beck & Katz 2001](#)). After discussing these issues, I show how they are actually a result of the same underlying problem: small-sample bias. In the former case we have too few realizations in the total sample to generate accurate estimates, while in the latter we have too few realizations in the unit sample to produce finite estimates.

As such, I propose a general ‘small sample’ solution, penalized maximum likelihood (PML), which can be applied to either of these cases. PML not only corrects the bias from small samples, as do current rare-event approaches, but it can also provide finite estimates for parameters even in instances of perfect separation. I show how this allows us to estimate a fixed-effects model where all units are retained, minimizing the risk of bias and improving efficiency. As such, PML provides a flexible solution to several types of challenges raised with rare-events data and, I argue, should be the preferred approach for modeling these data. Moreover, I detail how this strategy could be combined with group-selection techniques to estimate models of group, rather than unit, fixed effects. Group fixed-effects affords researchers greater flexibility than current methods which require them to assume either complete homogeneity (pooled) or heterogeneity (unit-fixed-effects).²¹ In so doing, it can also minimize the biases confronted when employing either of these ‘pure’ strategies by selecting a minimal set of groups (reducing the incidental-parameter bias of unit fixed-effects) necessary

²¹I focus exclusively on heterogeneity with respect to the intercept, omitting any discussion of unit variation in the estimation of model parameters as discussed in [Beck & Katz \(2007\)](#) and elsewhere.

to accurately model unit heterogeneity in the sample (avoiding the omitted-variable bias of pooled models).

In light of the theoretical importance of properly modeling unit heterogeneity, I (re-)analyze the relationship between GDP and civil war in [chapter 6](#). In current research, the pacific effect of development is largely unquestioned. It is widely regarded as the most robust finding in civil war research, with some concluding that the debate on this issue has been settled: poverty matters. These studies seem to confirm what we already know, after all, one need only “read the newspapers” to “see that the countries where there is conflict are far more likely to be poor” ([Collier 2008](#)). However, as in all things, correlation is not causation. There are many reasons why a state may be abundant in both poverty and civil war (e.g., weak institutions, inter-ethnic tensions, etc . . .), calling into question the direct relationship between the two. In such situations, a fixed-effects estimator is called for, allowing us to control these unobservables and examine the direct relationship between GDP and conflict.²² However, researchers in development and civil war have been hesitant to embrace and even advocate against the use of such methods, in part, it seems, because when utilized the relationship between GDP and civil war is no longer present. I argue that many of these concerns are unfounded (e.g., sample selection, nearly time-invariant regressors) and that fixed effects should be the preferred approach. As such, we need to take meaningfully the result that GDP and civil war do not appear to have a direct relationship. While this may be surprising to some, it actually confirms formal analysis of conflict which has found no basis for a casual explanation linking development to conflict ([Chassang & Padro-i Miquel 2009](#), [Fearon 2008](#)). As such, I argue it is time to reopen the debate linking development and poverty and consider alternative explanations for what remains an interesting empirical regularity.

Finally, in [chapter 7](#), I conclude by summarizing the main issues presented in the thesis. In particular, I discuss the need for future work unifying the two main issues addressed in the thesis: auto-dependence and unit-level heterogeneity. Being able to distinguish between these processes has significant theoretical consequences. For example, it is important to

²²See [Acemoglu et al. \(2008\)](#) for a celebrated example of this approach examining income and democracy.

understand whether civil conflict occurrence increases the risk of conflict or whether it is simply more prevalent amongst a certain subset of states (independent of any initial conflict). As I note, however, achieving this sort of unified model is no easy task. Furthermore, I recall what Beck has call “Stimson’s Law”: You can only solve one hard problem at a time, and solving requires ignoring lots of other problems.

2.0 (INTER)DEPENDENCE ACROSS TIME AND SPACE

After choosing the area we usually have no guidance beyond the widely verifiable fact that patches in close proximity are commonly more alike, as judged by the yield of crops, than those which are far apart.

— R.A. Fisher, 1935

Everything is related to everything else, but near things are more related than distant things.

— Waldo Tobler, 1970

That international events, decisions, outcomes, are not independent is obvious to even casual observers of global politics. One would be remiss to talk of the recent trouble in Northern Ireland – with ”a couple [of deaths] a year, and sectarianism [which] continues to plague society”¹ – without an appreciation of the Troubles which came before. Likewise, any analysis into the Arab Spring or the Great Recession without reference to the role of globalization would be nearly impossible and highly suspect. Despite this, much of the literature in International Relations either neglects the possible role of such dependence outright or fails to properly capture its true effect. In either respect, our inability to adequately account for, and generate theories pertaining to, such dependence risks biasing both our empirical findings and, more importantly, limits our understanding as to the true causes and consequences of these important events. As discussed in [chapter 1](#), this dependence is present in a host of important issues within International Relations and can be particularly difficult to account for when our questions pertain to binary and rare events, as they so often do.

¹Quote from ‘History trumps democracy’ in the Economist on March, 29 2014

Working under the incorrect assumption of spatial, temporal, or spatiotemporal independence results in overconfidence and inefficiency at best, and can also result in bias. Yet, incorrectly modeling this dependence can introduce bias into our estimates as well. Therefore, in this chapter I review and discuss the common strategies for dealing with temporal and spatial dependence in binary outcomes in International Relations.² Following this, I present an alternative simulation-based approach to estimate temporal and spatial dependence in non-linear data using maximum-simulated-likelihood (MSL) by recursive importance sampling (RIS) suggested by [Franzese *et al.* \(2014\)](#). As detailed in the remainder of the chapter, this approach has several advantages.

First, while current approaches – e.g., including a spatial lag as an exogenous right-hand-side regressor – suffer from simultaneity bias, thereby producing potentially wildly inflated estimates of the spatial lag and underestimates of the remaining model parameters. MSL-by-RIS accounts for the endogeneity of the spatial lag, providing accurate estimates of both the dependence and model parameters. Second, while the typical approaches for modeling temporal dependence are not as obviously flawed – in the sense that they tend not to bias the estimates of the betas – MSL-by-RIS is the only one which explicitly models the auto-regression in the outcome, permitting a range of ARMA specifications familiar to time-series analysis with interval data.³ More importantly, it allows us to estimate substantive conditional spatiotemporal effects, response-paths, and long-run-steady-state effects. Common strategies are ill-suited to generating these type of effect estimates, and as a consequence researchers have typically ignored substantive effect-estimates when estimating models with spatiotemporal dependence. Given that it is these effects that we are ultimately interested in,

²While much of the discussion on the current approaches would be germane to either logit or probit, I confine my attention to probit as it has been the most common estimator when modeling complex correlation structures in binary data. In part, this is because it is relatively easier to draw from an n-dimensional normal than an n-dimensional extreme-value. I return to this issue in [chapter 7](#) where I discuss how we might borrow insights from Exponential Random Graph Models (ERGM) to produce a spatial-logit and/or dynamic-logit estimator.

³Furthermore, MSL-by-RIS is the only approach which offers consistent estimates of *both* temporal and spatial dependence while maintaining a consistent assumption about the nature of the dependence in the outcome. In this case, that it operates through latent- y^* . Other potential approaches would be forced to defend that dependence in time occurs through *observed* y while dependence in space exists in latent- y^* , which seems theoretically confused.

I show how simulations can use the model parameters to estimate counter-factual substantive effects.

Improving our ability to account for spatial and spatiotemporal dependence is ultimately important *theoretically*. Recent history alone should motivate us to consider more closely the means by which events (e.g., crises) propagate globally. It is now evident, if it was not before, that such events are not independently determined, but instead the consequence of both domestic *and* international factors. Failing to incorporate the latter into our theories and subsequent analysis risks our ability to draw accurate conclusions over either. That is, until and unless we account for the interdependence of these outcomes, our understanding of the more traditional country-specific factors will be necessarily biased. These concerns are not simply academic, but as noted by [Allen & Gale \(2007\)](#), when discussing financial crises, “a full understanding of contagion is necessary before adequate policy responses can be designed” (28). The same could equally be said about almost any area of interest in International Relations. Appreciating this reality is only going to grow more imperative as the world continues to ‘shrink’ and the ties between states deepen and become more embedded. Consequently, it will only become more important that *International Relations* scholars have theories and models which can speak to these developments.

2.1 SPATIOTEMPORAL DEPENDENCE

Although dependence within or across units presents many of the same problems for credible estimation, substantially more work has explored the former concern. As such, I begin by focusing on dependence across time, before turning to cross-sectional dependence as well. Common strategies for modeling temporal dependence in binary data are discussed in [Jackman \(2000\)](#) and [Beck *et al.* \(2001\)](#), I borrow from and build upon those discussions here. As noted and defined in [chapter 1](#), the non-dynamic model is given by the following 2 equations:

$$y_{it}^* = \mathbf{X}_{it}\boldsymbol{\beta} + \epsilon_{it} \quad (2.1)$$

$$y_{it} = \begin{cases} 1, & \text{if } y_{it}^* > 0 \\ 0, & \text{if } y_{it}^* \leq 0 \end{cases} \quad (2.2)$$

which in the presence of serial dependence will, for starters, underestimate the standard errors, resulting in overconfident estimates of the parameter significance.

This model is rarely used with BTSCS data any longer, as researchers have increasingly recognized the presence of dependence in their data and appreciated the threats it poses for credible inference. In particular, the work of [Beck *et al.* \(1998\)](#) singularly inspired a shift in the way most researchers now approach temporal dependence in binary outcomes.⁴ [Beck *et al.* \(1998\)](#) recognized that binary-outcome data *is* grouped (discrete-time) duration data, and as such simple solutions are available to model the effect of time. Specifically, they advise researchers to model temporal trends by including a series of time-since-event dummies, which give the baseline probability (i.e., hazard) of failure (i.e., observing a one) in a period. [Carter & Signorino \(2010\)](#) propose a similar technique, with differences arising in the approach each takes to smoothing the hazard (splines versus cubic polynomials). In effect, these proposals both treat time as nuisance and include parameters – duration-specific fixed effects – which, once included, provide more reasonable estimates of the substantive parameters. While these approaches represent a substantial improvement over non-dynamic specifications, they do not capture the auto-dependence on a unit’s current propensity on its previous propensity, that is, temporal auto-regression in the outcomes.

Several strategies have been proposed which account for this more directly, and therefore allow for a meaningful interpretation of temporal dependence. In particular, two distinct theoretic explanations for auto-dependence can be examined ([Jackman 2000](#)). The first explanation is that past outcomes condition future behavior, that is, as a consequence

⁴[Beck \(2008\)](#) notes that this article is the most widely cited work published in the American Political Science Review in the last 25 years.

of experiencing the event the factors which determine future occurrences are altered. This explanation centers on the impact that the observed realization of the event has on shaping subsequent behavior; as such, models of this type are referred to as “observation drive.” Alternatively, we may suspect that dependence exists in the latent factors which produce the observed event. That is, an individuals propensity at time t is a function of their propensity at $t - 1$. Elsewhere, models of this type have been referred to as the ‘parameter driven.’

Two ‘observation-driven’ models are given by the restricted-transition and full-transition models (Beck *et al.* 2001). The restricted transition model is the simple and widely used, strategy of simply including the lagged-observed value of y on the right hand side of equation 2.1 as follows:

$$y_{it}^* = \mathbf{X}_{it}\boldsymbol{\beta} + \phi y_{i,t-1} + \epsilon_{it} \quad (2.3)$$

which specifies the current *propensity* of the event at time t to the *observed* realization of the event at time $t - 1$. While this captures the state dependence in the model, it is not analogous to familiar lagged- y models with continuous outcomes, but instead simply shifts the intercept of y_{it}^* by ϕ (Beck *et al.* 2001).

Jackman (2000) introduces a more-complete transition model, which permits distinct factors to explain state-switching and state-dependence:

$$\begin{aligned} y_{it}^* | (y_{i,t-1} = 0) &= \mathbf{X}_{it}\boldsymbol{\beta}_0 + \epsilon_{it} \\ y_{it}^* | (y_{i,t-1} = 1) &= \mathbf{X}_{it}\boldsymbol{\beta}_1 + \epsilon_{it} \end{aligned} \quad (2.4)$$

which can be combined into a single conditional model:

$$y_{it}^* | (y_{i,t-1}) = \mathbf{X}_{it}\boldsymbol{\beta}_0 + \phi y_{i,t-1} \mathbf{X}_{it}\boldsymbol{\alpha} + \epsilon_{it} \quad (2.5)$$

where $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0 + \boldsymbol{\alpha}$, that is, the difference between parameter vectors for the distinct transition probabilities. While it enjoys some advantages – namely, it allows us to specify and

evaluate different models for event onset and event dependence simply in a single equation – it has not achieved widespread use.⁵

Alternatively, we can model temporal dependence in the *latent* propensities using ‘parameter driven’ approaches. Which, assuming an AR(1) process, is modeled as:

$$y_{it}^* = \mathbf{X}_{it}\boldsymbol{\beta} + \phi y_{i,t-1}^* + \epsilon_{it} \quad (2.6)$$

which is directly analogous to lagged-y models of continuous outcomes. This means that it can flexibly permit a variety of ARMA specifications familiar in dynamic time-series models. Furthermore, it allows us to assess theoretical claims about the persistence of unobserved (or unmeasured) variables across time. That is, a units propensity to experience an event conditional on its prior propensity. Accounting for this relationship also gives a natural interpretation to the substantive parameters in a way that none of the other models can easily achieve. Namely, the estimated $\boldsymbol{\beta}$ represent the immediate (next period) effect on the latent outcome, with the total effect on the latent variable given by the familiar $\frac{\beta}{1-\phi}$.⁶

An additional benefit of this approach, as will be shown, is that the same strategy can be employed to model cross-sectional (i.e., spatial) dependence as well. Unlike with temporal dependence, there are no simple evasions of cross-sectional dependence. Current strategies most often include either *i*) estimating a standard (non-spatial) probit or logit model or *ii*) introducing a spatial-lag on the right-hand-side of a model as if it were exogenous. The former strategy is obviously wrong, resulting in biased estimates from the omission of the spatial lag. While better intentioned, the latter is also wrong as it fails to account for the endogeneity (i.e., simultaneity) in the spatial lag.⁷ As such, estimates from these naïve spatial models will also be biased. Furthermore, placing the observed binary outcomes of other units simultaneously is not algebraically consistent, as it can only logically operate

⁵Perhaps this is because both of the papers (Beck *et al.* 2001, Jackman 2000) discussing this approach remain unpublished.

⁶Elsewhere, Beck (2001) has also proposed a “strain relief” estimator which is akin to the error correction model for continuous outcomes, replacing the lagged-latent $y_{i,t-1}^*$ with $y_{i,t-1}^* - \gamma y_{i,t-1}$.

⁷Time-lagging the spatial lag typically offers little relief from this concern. As Beck *et al.* (2006) note time-lagging the spatial lag can evade the simultaneity bias but the assumptions required for this to be valid rarely hold (see Franzese *et al.* 2014).

through the latent variables or errors (Heckman 1978). Equation 2.6 shows how we can do this, rewriting it for spatial dependence:

$$y_{it}^* = \mathbf{X}_{it}\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{y}^* + \epsilon_{it} \quad (2.7)$$

where \mathbf{W} is a weights matrix identifying connections between units and \mathbf{y}^* is a vector of the latent outcomes, which, taken together, is the spatial lag $\mathbf{W}\mathbf{y}^*$. In this formulation the propensity for unit i is conditional on the propensity of unit $j \neq i$ up to some weighting parameter w_{ij} . As such, we can see how the model for latent- y^* can be written to accommodate both spatial and temporal dependence. Yet, the largest complication remains: estimation. Despite their desirable properties, little work has pursued this approach given that estimating these models has been described as “notoriously difficult” given the “formidable” expression of the joint probabilities which poses a “ferocious maximization problem” Jackman (2000). I briefly explain the econometric challenges that are raised in estimating these models, before raising a simulation-assisted solution.

With time-series cross-sectional (TSCS) data - both temporal and cross-sectional dependence – the spatiotemporal probit model takes the structural form:

$$\mathbf{y}^* = \rho\mathbf{W}\mathbf{y}^* + \phi\mathbf{L}\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.8)$$

Which, written in reduced form is given as:

$$\mathbf{y}^* = (\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \text{ with } \mathbf{u} = (\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}\boldsymbol{\epsilon} \quad (2.9)$$

Latent-variable y^* links to observed y through the standard measurement equation:

$$y_{it} = \begin{cases} 1, & \text{if } y_{it}^* > 0 \Rightarrow [(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}\boldsymbol{\epsilon}]_{it} > -[(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}\mathbf{X}\boldsymbol{\beta}]_{it} \\ 0, & \text{if } y_{it}^* \leq 0 \Rightarrow [(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}\boldsymbol{\epsilon}]_{it} \leq -[(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}\mathbf{X}\boldsymbol{\beta}]_{it} \end{cases} \quad (2.10)$$

Where y_{it} is an observed binary outcome, \mathbf{X} is a $k \times NT$ matrix of covariates, and $(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}$ is the spatiotemporal multiplier consisting of an $NT \times NT$ identity matrix, \mathbf{I} , an $NT \times NT$ spatial-weights matrix \mathbf{W} , and an $NT \times NT$ time-shift matrix \mathbf{L} . More simply, each element $w_{it,jt}$ of \mathbf{W} identifies whether (or to what extent) units i and j are related spatially at time t . While, matrix \mathbf{L} is binary, with ones – for a single-period lag – indicating the previous period for the same cross-sectional unit (it in the row and $i,t-1$ in the column).⁸

The probability that the it^{th} observation is one is calculated as follows:

$$\begin{aligned} p(y_{it} = 1 | \mathbf{X}_{it}) &= p([\mathbf{I} - \phi\mathbf{L} - \rho\mathbf{W}]^{-1}\mathbf{X}\boldsymbol{\beta}]_{it} + [\mathbf{I} - \phi\mathbf{L} - \rho\mathbf{W}]^{-1}\boldsymbol{\epsilon}]_{it} > 0) \\ &= p(u_{it} < [(\mathbf{I} - \phi\mathbf{L} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_{it} / \sigma_{it}) \\ &= \Phi_{it}\{[(\mathbf{I} - \phi\mathbf{L} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_{it} / \sigma_{it}\} \end{aligned} \quad (2.11)$$

Therefore, as in standard probit (see 1.4, a cumulative-normal distribution, $\Phi_{it}\{\cdot\}$, gives the probability that that systematic component, $[(\mathbf{I} - \phi\mathbf{L} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_{it} / \sigma_{it}$, exceeds the stochastic component, u_{it} . However the $\Phi_{it}\{\cdot\}$ here is the it^{th} marginal probability from the n -dimensional cumulative-normal evaluated at the n cutpoints, $[(\mathbf{I} - \phi\mathbf{L} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_{it} / \sigma_{it}$, because the interdependence in latent- y_{it}^* in spatiotemporal probit induces nonsphericity in the stochastic components \mathbf{u} . Specifically, \mathbf{u} is distributed n -dimensional multivariate normal with mean 0 and variance-covariance $[(\mathbf{I} - \phi\mathbf{L} - \rho\mathbf{W})'(\mathbf{I} - \phi\mathbf{L} - \rho\mathbf{W})]^{-1}$. Computing these probabilities is intense as one must read the probability that $[(\mathbf{I} - \phi\mathbf{L} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_{it} / \sigma_{it}$ exceeds u_{it} from the it^{th} marginal distribution of the multivariate cumulative-normal $\Phi_{it}\{\cdot\}$, which requires integrating that joint distribution over *all* n dimensions. Also, because σ_{it} is not constant – the it, it^{th} element of the variance-covariance – interdependence also induces heteroscedasticity. This heteroscedasticity and, more crucially, the interdependence (i.e., the non-independence) for the u render standard probit inappropriate and generate the computation intensity. That is, because the outcomes are interdependent, their joint distribution is not the product of

⁸For example, in the single-period lag instance, the \mathbf{L} matrix would be all zeroes except for ones along the diagonal of the lower block first-minor.

the n univariate marginal distributions, instead maximizing the log of one non-separable n -dimensional distribution.

2.2 MSL-BY-RIS: ESTIMATION

High dimensional integrals, like the one encountered here, are difficult to calculate numerically, but can be well approximated via simulation (Train 2009). Drawing upon the strategy first proposed by Beron *et al.* (2003) and Beron & Vijverberg (2004), Franzese *et al.* (2014) offer a maximum simulated-likelihood (MSL) by recursive-importance-sampling (RIS) strategy to estimate spatiotemporal qualitative dependent variable models. In this section I briefly describe the technical details of MSL-by-RIS detail its performance against standard estimators, outline a related simulation-based strategy for calculating substantive effects, and conclude with a discussion of its applicability to rare events.

RIS approximates densities which are difficult to calculate analytically, such as the cumulative multivariate normal distribution introduced by interdependence in spatiotemporal probit. Given our inability to draw from $f(\mathbf{x})$ (i.e., the target density), we can instead approximate the sought probabilities by taking repeated draws from a better known density with the same support $g(\mathbf{x})$ (i.e., the proposal density) and weighing them by $f(\mathbf{x})/g(\mathbf{x})$ (Train 2009). When repeated R many times, the weighted draws are equivalent to draws from original target density, as the CDF of the weighted draws of $g(\mathbf{x})$ is the same as the CDF of draws from $f(\mathbf{x})$ itself.

That is, to approximate an n -dimensional cumulative multivariate-normal:

$$p = \int_{-\infty}^{\mathbf{x}_0} f_n(\mathbf{x}) d\mathbf{x} \quad (2.12)$$

where $f_n(\mathbf{x})$ is the density and $[-\infty, \mathbf{x}_0]$ the interval over which one wants to integrate, one chooses n -dimensional sampling-distribution with well-known properties, and defines a truncation of this distribution with support over the same interval as $g_n^c(\mathbf{x})$. Multiplying

and dividing the right-hand-side of the integral one wishes to calculate (e.g., [Equation 2.12](#)) by this density:

$$p = \int_{-\infty}^{\mathbf{x}_0} \frac{f_n(\mathbf{x})}{g_n^c(\mathbf{x})} g_n^c(\mathbf{x}) d\mathbf{x} \quad (2.13)$$

This integral is a mean which gives the probability sought, p , as the mean of $\frac{f_n(\mathbf{x})}{g_n^c(\mathbf{x})}$, which can be estimated using a sample of R draws of the vector \mathbf{x} from the importance distribution:

$$p = E \left[\frac{f_n(\mathbf{x})}{g_n^c(\mathbf{x})} \right] \approx \frac{1}{R} \sum_{r=1}^R \frac{f_n(\tilde{\mathbf{x}})}{g_n^c(\tilde{\mathbf{x}})} \equiv \hat{p} \quad (2.14)$$

With RIS, we draw \mathbf{x} from a truncated multivariate normal and calculate $\frac{f_n(\mathbf{x})}{g_n^c(\mathbf{x})}$.

While this would serve in suffice in standard probit with independent errors, we still haven't confronted the issue of interdependent errors, as the numerator in [equation 2.14](#) is still a single n -dimensional cumulative-normal. However, given that the variance-covariance matrix, $\Sigma = [(\mathbf{I} - \phi\mathbf{L} - \rho\mathbf{W})'(\mathbf{I} - \phi\mathbf{L} - \rho\mathbf{W})]^{-1}$, is positive-definite we can use Cholesky factorization to decompose the NT correlated probabilities as NT independent components. That is, a decomposition exists such that $\Sigma^{-1} = \mathbf{C}'\mathbf{C}'$ with \mathbf{C} an upper-triangular matrix and $\boldsymbol{\eta} \equiv \mathbf{C}\mathbf{u}$ given n independent standard-normal variables, η . Let $\mathbf{B} \equiv \mathbf{C}^{-1}$ and substitute $\mathbf{u} = \mathbf{C}^{-1}\boldsymbol{\eta} \equiv \mathbf{B}\boldsymbol{\eta}$. From this we can determine the sought probability, $p(\mathbf{u} < \mathbf{v}) = p(\mathbf{B}\boldsymbol{\eta} < \mathbf{v})$ by evaluating the CDF at the implied upper bounds, which are determined recursively starting with the last observation and then multiplying these probabilities:

$$\begin{aligned} \eta_n &< b_{nn}^{-1}v_n \equiv \eta_{n0} \\ \eta_j &< b_{jj}^{-1} \left(v_j - \sum_{i=j+1}^n b_{ji}\eta_i \right) \equiv \eta_{j0}(v_j, \eta_{j+1}, \dots, \eta_n) \equiv \eta_{j0} \end{aligned} \quad (2.15)$$

The probability of having observed a sample of ones and zeros can now be found by evaluating the univariate CDF at each of these bounds and then multiplying those probabilities:

$\prod_{j=1}^n p_j = \prod_{j=1}^n \Phi(v_j)$. Repeating R times and averaging gives the maximum simulated likelihood \hat{l} which, given enough runs of R , is asymptotically equivalent to maximum likelihood. As in ML then, it provides consistent estimates of the parameters.

The performance of MSL-by-RIS is examined via monte carlo analysis in [Franzese *et al.* \(2014\)](#), I highlight some of our more significant findings here.⁹ The data-generating process for the Monte Carlos closely follows that of [Beron & Vijverberg \(2004\)](#), but expands to include a time dimension. In particular, the DGP takes the form:

$$\mathbf{y}^* = (\mathbf{I}_n - \rho \mathbf{W} - \phi \mathbf{L})^{-1}(\beta_0 + \beta_1 \mathbf{x} + \boldsymbol{\epsilon}), \boldsymbol{\epsilon} \sim N(0, 1) \quad (2.16)$$

with the measurement equation linking y^* to y . For \mathbf{W} , a row-standardized binary-contiguity matrix of the 50 U.S. states is used. Data for each unit is generated for 20 periods, giving a sample size of 1000 in each of the reported monte carlo experiments. In these experiments $\beta_0 = 1.5$ and $\beta_1 = 3.0$ are held fixed, while $\rho = 0.10, 0.25$ and $\phi = 0.3, 0.5$ vary to represent difference levels of dependence. Lastly, \mathbf{x}_0 is drawn from a standard uniform distribution on the interval $[-1, 2]$, resulting in an expected value of 0.5 and a variance of roughly 2. The results are presented in [Table 2.1](#), with 100 trials for each experiment and $R=100$, and provides support for the estimator. That is, the estimates are all quite accurate when the data follow a spatiotemporal DGP.

⁹Many results will not be presented or discussed here. If, for example, the reader is interested in how MSL-by-RIS performs as compared to traditional strategies for temporal dependence (e.g., time-since-event counters, regime-switching models, etc...) they should refer to the article.

Table 2.1: Simulation Results for MSL-by-RIS Coeff. Est.

	$\beta_0=-1.5$	$\beta_1=3.0$	ϕ	ρ
Experiment #1: $\rho=0.10, \phi=0.30$				
Coeff. Est	-1.467	2.962	0.092	0.276
RMSE	0.104	0.173	0.043	0.032
Std Dev	0.101	0.169	0.042	0.021
SE	0.129	0.231	0.045	0.025
Overconfidence	0.784	0.730	0.938	0.840
Experiment #2: $\rho=0.10, \phi=0.50$				
Coeff. Est	-1.385	2.797	0.097	0.464
RMSE	0.150	0.273	0.034	0.042
Std Dev	0.095	0.183	0.034	0.021
SE	0.105	0.199	0.042	0.020
Overconfidence	0.906	0.920	0.802	1.022
Experiment #3: $\rho=0.25, \phi=0.30$				
Coeff. Est	-1.450	2.922	0.224	0.280
RMSE	0.117	0.197	0.047	0.028
Std Dev	0.106	0.181	0.038	0.020
SE	0.119	0.212	0.046	0.024
Overconfidence	0.884	0.854	0.834	0.833
Experiment #4: $\rho=0.25, \phi=0.5$				
Coeff. Est	-1.363	2.752	0.241	0.471
RMSE	0.172	0.322	0.034	0.035
Std Dev	0.104	0.205	0.033	0.019
SE	0.107	0.202	0.031	0.020
Overconfidence	0.969	1.013	1.043	0.946

2.3 MSL-BY-RIS: EFFECTS

Ultimately, however, we are not interested in parameter estimates as such, but the effects $\frac{\delta p(y_i=1)}{\delta x_i}$ they suggest. Estimating these effects is complicated in the presence of spatial or spatiotemporal interdependence as even *within*-unit counterfactuals – e.g., x_i on y_i – involve feedback from i through other units $j \neq i$ back to i . Furthermore, with dependent data we are often interested in cross-unit effects as well, up to and including, counterfactual outcomes in other units $\frac{\Delta p(y_i=1)}{\Delta y_j}$. As in estimation, interdependence complicates our ability to generate such conditional effects. While we could estimate these in a manner analogous to

that used in estimation – using RIS to approximate the conditional probability sought – this would be extremely computationally burdensome and neglect more expedient alternatives.

[Franzese *et al.* \(2014\)](#) show how these counterfactuals can be calculated, using the definition of conditional probability they are simply:

$$\begin{aligned} p[y_{i,t+s} = 1 | y_{j,t} = 1; \mathbf{X}, \mathbf{W}, \mathbf{L}] - p[y_{i,t+s} = 1 | y_{j,t} = 0; \mathbf{X}, \mathbf{W}, \mathbf{L}] \\ = \frac{p[y_{i,t+s} = 1, y_{j,t} = 1 | \mathbf{X}, \mathbf{W}, \mathbf{L}]}{p[y_{j,t} = 1 | \mathbf{X}, \mathbf{W}, \mathbf{L}]} - \frac{p[y_{i,t+s} = 1, y_{j,t} = 0 | \mathbf{X}, \mathbf{W}, \mathbf{L}]}{p[y_{j,t} = 0 | \mathbf{X}, \mathbf{W}, \mathbf{L}]} \quad (2.17) \end{aligned}$$

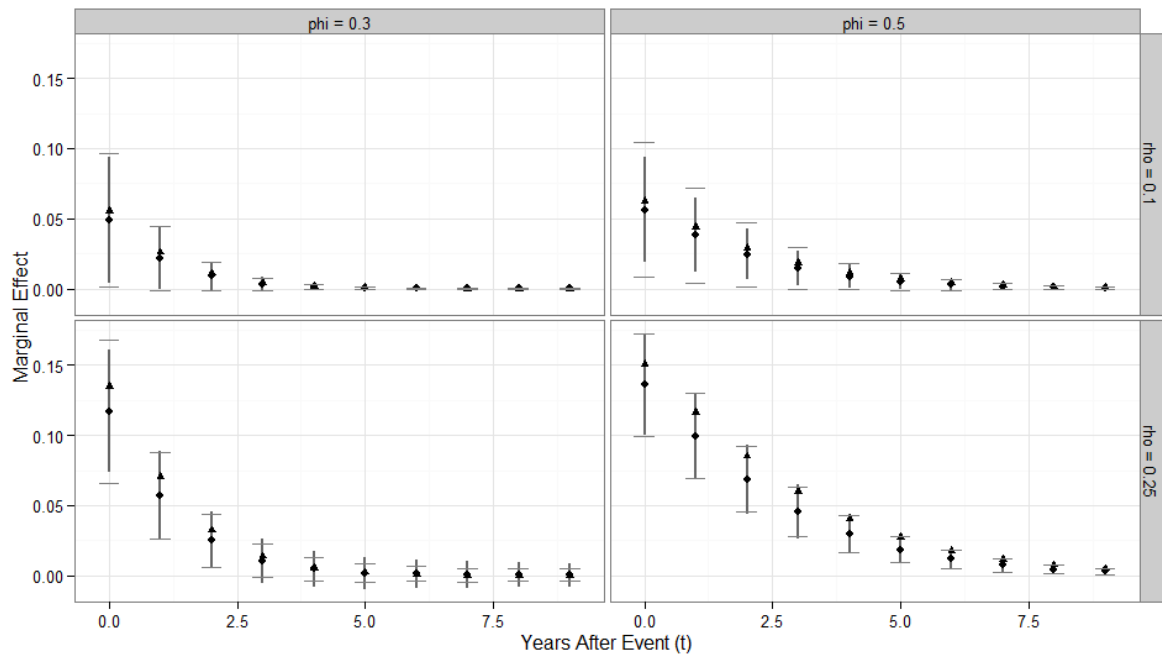
With the estimates of ρ , ϕ , and β , one calculates the CDFs for the two univariate (denominators) and two multivariate-normal (numerators) distributions to produce an estimated spatiotemporal responses path. Standard errors for these point estimates are calculated via parametric bootstrap: repeat the procedure for each of the many draws of the parameter estimates from their estimated joint distribution, and average the standard deviation across draws to give the estimate and its standard error.

The accuracy of this approach is shown in [Figure 2.1](#). Selecting a specific pair of units i and j (Alabama and Mississippi) we first calculate the true effect, $p[y_{i,t+s} = 1 | y_{j,t} = 1; \mathbf{X}, \mathbf{W}, \mathbf{L}] - p[y_{i,t+s} = 1 | y_{j,t} = 0; \mathbf{X}, \mathbf{W}, \mathbf{L}]$, assuming \mathbf{x} equals its last sample-values. Then, using the parameter estimates from the monte carlos, we calculate the effects using the spatiotemporal-lag probit model as just described. [Figure 2.1](#) provides information about: *i*) the bias in the response paths (comparing the truth triangles to the average-estimate diamonds); *ii*) the efficiency of these estimates (the vertical lines indicate the standard deviation across trials); and *iii*) the accuracy of our uncertainty estimates (the horizontal ticks indicate the estimated standard errors).¹⁰ By and large, we observe that the estimated response-paths are very accurate.

While this strategy suffices with lower dimensions (i.e., the number of counterfactual conditions), it becomes computationally burdensome (or impossible) with higher-order

¹⁰That is, comparing the vertical lines to the horizontal lines indicates the (over)confidence of the estimators standard error estimates.

Figure 2.1: Accuracy of RIS for Long-Run Response Paths



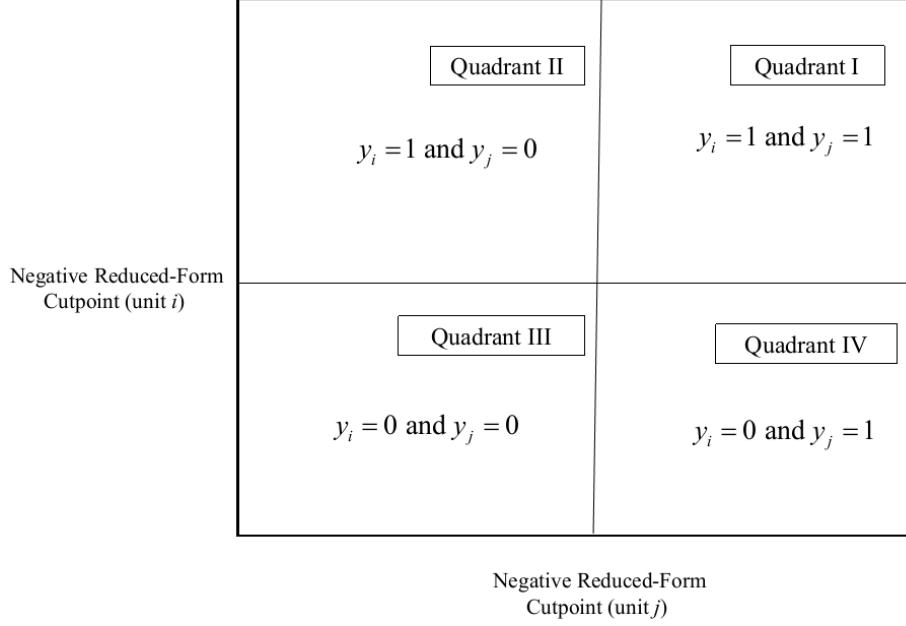
◆ Estimated response-paths; ▲ True response-paths
Horizontal bars indicate 95% confidence intervals using the estimated standard errors (within trials);
Vertical lines indicate 95% intervals using the actual standard deviation of the effect estimates (across trials).

dimensions. In these instances an alternative strategy may be preferable: brute-force simulation. First proposed by [Hays \(2009\)](#), one can simply draw from the disturbances and coefficients (using the sample estimates as the mean for the draws), pre-multiply each by the spatiotemporal filter to generate reduced-form disturbances and reduced-form cutpoints (for some pre-selected and fixed counterfactual $\mathbf{X}_1 \mathbf{X}_0$), and generate a vector of hypothetical realizations of y_1 and y_0 . In other words, we can use the estimated model to generate counterfactual probabilities of interest.

As with the last strategy, this technique can be used to generate probabilities and effects conditional on observed outcomes in others units, for example $p[y_{i,t+s} = 1 | y_{j,t} = 1; \mathbf{X}, \mathbf{W}, \mathbf{L}]$ and $p[y_{i,t+s} = 1 | y_{j,t} = 0; \mathbf{X}, \mathbf{W}, \mathbf{L}]$. As I show in [chapter 3](#) and ?? for many applications these kind of counterfactuals are often more substantively interesting than an average effect. The probabilities are ratios of the quadrant counts from a 2-dimensional graph where the axes represent the negative of the reduced-form cutpoints for units i and j , the i th and j th elements of the vector $-(\mathbf{I}_n - \rho \mathbf{W})^{-1} \mathbf{X} \hat{\boldsymbol{\beta}}$. [Figure 2.2](#) helps to illustrate how we can use the parameter estimates to generate simulated realizations of the counter-factual outcome. A single draw from the reduced-form disturbances for units i and j , the i th and j th elements of vector $-(\mathbf{I}_n - \rho \mathbf{W})^{-1} \boldsymbol{\epsilon}$ identifies an x-y coordinate (a point) located in one of the quadrants. If this point is in quadrant I, for example, the reduced-form disturbances for both i and j are above the negative of their respective reduced-form cutpoints and $y_i = 1$ and $y_j = 1$. Conditional relative frequencies, which are ratios of quadrant counts, provide estimates for the conditional probabilities of interest. Specifically, the probabilities are estimated by:

$$\begin{aligned} \Pr[y_i = 1 | X, \mathbf{W}, y_j = 1] \\ \Pr[y_i = 1 | X, \mathbf{W}, y_j = 0] \end{aligned} \tag{2.18}$$

Figure 2.2: Estimating Spatial Effects via Simulation



As before, to reflect the uncertainty about the model's parameters (i.e., the cutpoints), we can draw from the estimated sampling distribution for the parameters. Each draw generates a new set of reduced-form cutpoints which leads to a new set of conditional probabilities. The results produced using this method are equivalent to those obtained using the other technique and will occasionally may prove more computationally efficient (and/or feasible) in higher dimensions.¹¹

With either proposed method for estimating effects, we are able to generate long-run spatiotemporal response paths (as in Figure 2.1). That is, we can assess to the effect of a shock in t many periods into the future. Asking, for example, what the impact of unit j experiencing an event at t has on unit i 's probability of experiencing an event at $t + 5$? This is one of the chief benefits of MSL-by-RIS as no existing strategy and easily or straight-forwardly calculate these dynamic response paths for binary data. Furthermore, by

¹¹Results detailing the efficacy of this brute-force simulation strategy available upon request.

conditioning on several outcomes we can generate different counter-factual *histories* for a unit (or several units). This allows researchers to examine the differences in current probabilities that obtain from several potential historical paths ranging from never having experienced the event (i.e., all zeroes) to incidence in every years (i.e., all ones).

2.4 DISCUSSION

That the outcomes for one unit(-time) depend on the outcomes of other unit(-time)s seems obvious. This is only made more likely when the population is small, well-defined, and placed in a system where interactions occur often and repeatedly, such as international affairs. Yet, International Relations scholars have only recently began to appreciate the impact this has only almost all over our theoretical questions. As I noted at the outset, this is particularly true for binary outcomes, where interstate and civil wars, economic shocks and sanctions, are all clearly interrelated. In this chapter, I described how dependency can manifest in binary-TSCS data and presented a simulation based strategy for estimating this dependence in terms of latent- y^* . Monte Carlo experiments have revealed the efficacy of this estimation strategy in obtaining accurate estimates for both the dependency parameters and the substantive parameters of interest. Furthermore, I have noted how, once obtained, these parameters can be used to estimate counter-factual conditional effects using related simulation methods. Thereby allowing for the estimation of the change to short- and long-run outcome probabilities associated with a hypothetical shift to some X , up to and including changes in the same-period outcome of other units.

One question which remains to be explored is the efficacy of these strategies for explicitly *rare* events, as we so often find in International Relations. [Franzese *et al.* \(2014\)](#) find that MSL-by-RIS exhibits the well-known small sample bias in maximum likelihood estimates of spatial-lag models. That is, likelihood estimates of the dependence parameters (ρ and ϕ) are typically downwardly biased, which induces inflated estimates of the remaining parameter estimates (β). While a similar bias is found in spatial-ML estimates for continuous outcomes,

it is a more pernicious problem with binary outcomes as indicated here. With rare events the challenge may be even greater as the rate of switching – which is required to update ρ and ϕ – is lower. In short, rare events only further exacerbate the “small sample” problem and can persist in even relatively large samples if the event count remains low.¹² This likely implies when estimating models of rare events the dependency parameter estimates will be further attenuated. Given that many of the applications of this method will be into areas with rare events, it is important to get a handle on the extent of this bias. Therefore, in the future I plan to run experiments varying the sparseness of the data when estimating MSL-by-RIS to assess the extent of this downward bias with rare events.

Furthermore, there are deeper theoretical questions about the nature and dynamics of interdependence which will require more computationally efficient estimation techniques. While showing that outcomes are interdependent is an important first step, it neglects the other ways in which states likely respond to these anticipated spillovers. These actions could, in turn, influence the likelihood of witness both spillovers to that state *and*, more interestingly, whether we observe an event in the originating states at all. As a toy example, anticipating a potential crisis in state i , state j undertakes some intervention intended to reduce this possibility, motivated to do so because the ties between state i and j risk contagion in the event of a crisis. Now the event outcome (e.g., crisis) in states i and j are not just dependent upon one another, but also the related policy outcome(s) (e.g., intervention) of these states as well:

$$\begin{aligned} &\Pr(\text{crisis}_i = 1 | \text{crisis}_j, \text{intervention}_i, \text{intervention}_j) \\ &\Pr(\text{crisis}_j = 1 | \text{crisis}_i, \text{intervention}_i, \text{intervention}_j) \\ &\Pr(\text{intervention}_{ij} = 1 | \text{crisis}_j, \text{crisis}_i, \text{intervention}_{ji}) \\ &\Pr(\text{intervention}_{ji} = 1 | \text{crisis}_j, \text{crisis}_i, \text{intervention}_{ij}) \end{aligned}$$

This provides a much richer theoretical account about the possible strategic role states play in the co-determination of international events. Furthermore, such analysis would provide

¹²This poses problems for ML-estimation in general as discussed in [chapter 5](#)

more insights into the incentives of states when determine to take some interventions and not others. However, this also greatly expands the analysis as the intervention outcome is (in this formulation) dyadic, that is, j intervening into i , etc... Furthermore, absent some arbitrarily imposed sample constraint, every possible dyadic combination is able to intervene, with the decision to do so a function of all other choices. This is significantly more computationally burdensome – the estimation sample is no longer N but N -choose-2 – and as such improvements on or alternatives to MSL-by-RIS will need to be pursued.¹³

¹³([Wilhelm & de Matos 2013](#)) new R package ‘spatialprobit’ implements an MCMC approach using Gibbs sampling, the Bayesian analog to the method discussed here, which appears to possess considerable processing advantages and as such may prove useful for these purposes.

3.0 THE DYNAMICS OF CIVIL WAR

The past is prophetic in that it asserts loudly that wars are poor chisels for carving out peaceful tomorrows.

— Martin Luther King, Jr., 1967

The principal cause of war is war itself.

— C. Wright Mills, 1959

In the last 65 years there have been more than 300 civil conflicts.¹ The stark consequences of these conflicts – displaced populations, increased crime, environmental degradation, death tolls in the millions, economic destruction in the billions, etc – have generated an immense research agenda attempting to better understand and explain the factors which cause these events. One key finding of this research is that civil wars themselves foster the conditions that increase the risk of further conflict. That is, the causes and consequences of civil war are one in the same: economic instability, political uncertainty, social unrest, a mobilized and divide populace. As such, the ultimate consequence of civil war is often more war, with states caught in a recursive sequence of conflict (i.e., the ‘conflict trap’). Given that the majority of conflicts take place within these countries, understanding the nature of conflict recurrence is essential for explaining civil war occurrence.

A substantial literature examines the persistence of conflict, whereby an initial conflict significantly increases the risk of future fighting (Collier *et al.* 2003; 2008, Hegre *et al.* 2011, Walters 2004). This research has focused on characteristics of the country (e.g., economic

¹Defined as violent incidents between the government and organized opposition within a state that results in at least 25 battle-related deaths (Themnér & Wallensteen 2013)

underdevelopment, natural resources) and the conflict (e.g., duration, issue, outcome) which make recidivism more likely. These factors, it is argued, explain why we observe the clustering of conflicts within a set of countries. However, these conflicts not only cluster *within* but *across* countries, with neighboring states frequently undergoing conflict concurrently. Moreover, many of the countries used to illustrate the persistence of conflict (e.g., Burundi, Rwanda, the Democratic Republic of Congo), are the same states used as anecdotes in the literature on regional conflict contagion, wherein scholars argue that neighboring conflicts increase the risk of civil war (Buhaug & Gleditsch 2008, Diehl 1991, Gleditsch 2002, Ward & Gleditsch 2002). In fact, what we observe is not just states, but entire *regions* mired in recurrent conflict. Yet, we lack a clear understanding of the process by which this emerges: is it a country's own history with conflict (i.e., persistence), the outbreak of war in a neighboring state (i.e., contagion), or both, that makes these protracted conflicts more likely?

I argue that any attempt to understand the means by which conflict breeds conflict must fully consider both of these dimensions, that is, the dependence of civil war across both time and space. Failure to do so risks misunderstanding the process by which these cycles or patterns of conflict emerge, erroneously favoring one or the other. Specifically, I argue that the contagion of civil conflict is a *cause* of its persistence, and vice-versa. Each states risk of conflict influences, and is influenced by, their neighbors risk of conflict, creating an additional cycle through which civil war persists. In this way the contagion of conflict makes persistence more likely, and, in turn, longer conflicts make contagion more likely. As such, the so-called 'conflict-trap' is, in part, a regional phenomenon, wherein each individual states sporadic episodes of conflict are augmented by those of their neighbors. This feedback produces conflict prone regions – or, conversely, peace prone regions – with persistently higher risks of conflict than would be expected from any single state alone, explaining why peace has been so difficult to obtain in particular states and regions (e.g., the Great Lakes region of Africa, West Africa, parts of Southeast Asia). In sum, it is both the dependence of conflict within a country *and* the interdependence of conflicts across countries that makes civil war recurrence so common.

Little work in the civil war literature has modeled the dependence of civil conflict in time or space in a manner which can provide insights into these dynamics, and no work allows us to directly discriminate between the two. Instead, I utilize the approach introduced in [chapter 2](#) which uses discrete-choice simulation to produce consistent estimates of *both* the dependence of conflict across time and space. This approach has several advantages over common alternatives in the literature. First, it directly models the temporal-autoregressive process, that is, the auto-dependence of a states conflict propensity on its prior conflict propensity. Unlike alternatives (e.g., peace years with splines, cubic polynomials), this allows us to model the dynamic responses we are interested in when examining conflict recurrence. Stated more simply, it allows us to directly measure the persistence of civil war risk. Second, it allows us to estimate the simultaneous spatial effects of conflict, whereas common approaches (e.g., including a spatial lag of neighboring conflict as a regressor) are necessarily biased. As such, this is the first work in the conflict literature to offer consistent estimates of the spatial effect of civil war. Finally, it permits more direct comparisons between the effects of previous conflict and neighboring conflict, allowing us to decompose effects into short- and long-run, spatial and non-spatial, to get a more complete understanding of the dynamics of civil war. In all, it allows us to better understand whether, how, and the extent to which, conflict begets conflict.

More than most, this issue has clear and important policy consequences. That conflict persists across time and spreads amongst neighbors is widely accepted as fact by the policy community and popular media alike. These beliefs drive numerous large-scale policy decisions, from the creation of the UN Peacebuilding Commission, addressing conflict recurrence, to the hundreds of millions in assistance given to countries near conflict zones, most recently Jordan, Turkey, and Lebanon in response to the Syrian civil war. As such, the challenge is not in convincing policy makers *that* these relationships exist, but rather in ensuring that we have the *correct* understanding of these dynamics. [Suhrke & Samset \(2007\)](#) chronicle how previous estimates on the risk of conflict recurrence quickly spread, became conventional wisdom, and shaped policy debates, only to later be significantly amended by these same authors (the work of Collier and his co-authors). In addition, they highlight two problems of providing effects estimates for ‘the typical country,’ arguing that: *i*) different

interpretations of ‘typical’ produce significantly different estimates, and *ii*) it is difficult to know what these estimates mean in any particular context. This paper attempts to redress these shortcomings, providing consistent estimates of spatiotemporal dependence in conflict, and a strategy to obtain country-specific effects estimates for substantively important cases via simulation.

I proceed with the analysis as follows. In the first section, I discuss the relevant literatures on the persistence and spread of civil conflict. Within this, I detail my argument on how these dynamics are necessarily related and therefore require a unified model. Following that, I evaluate the dependence of civil conflict using a spatiotemporal probit model of conflict incidence using maximum simulated-likelihood with recursive-importance-sampling. In addition, I show how conventional estimation strategies – ones with no or naïve measures of the spatial and/or temporal dependence – can produce substantially different results. After estimating these models, I use simulation to produce counter-factual estimates of the short and long-run effect that a civil war in one country has its risk of later conflict and the risk of conflict in neighboring states. This strategy also allows me to explore the cumulative effect of specific conflict histories, which, to my knowledge, do not exist in the current literature. Finally, I discuss the implications of these findings for our understanding of civil war as a recursive process and conclude.

3.1 A REGIONAL THEORY OF CONFLICT TRAPS

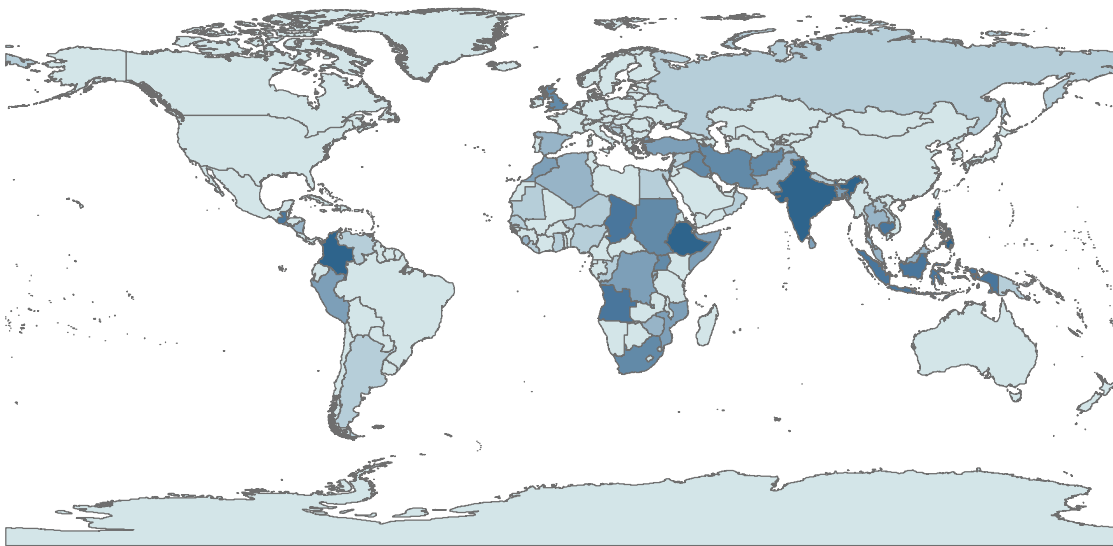
Since World War II, civil war has been the most prevalent form of armed conflict.² In any given year, close to 30 countries are engaged in active civil conflict. However, as illustrated in Figure 3.1, these events are not evenly distributed.³ Instead, we see that particular countries experience far greater rates of civil conflict, with as few as 15 countries accounting for more than *half* of the conflict incidents since 1950. If anything, recent patterns

²See Buhaug, et al. (2007) for a detailed discussion on the global trends in armed conflict.

³Data come from UCDP/PRIO (detailed below), with the lightest blue indicating no conflict episodes and the darkest blue indicating frequent incidents of civil conflict.

suggest that these episodes of civil conflict are becoming increasingly concentrated. An expanding number of states are effectively at no risk of conflict, yet the probability of civil war for the remaining pool of at-risk states continues to grow (Collier *et al.* 2003). What explains these differences, that is, why do some countries experience civil conflict with such regularity?

Figure 3.1: Global Incidence of Civil Conflict, 1950-2000



Note: The lightest blue indicates no conflict with successively darker shades denoting a greater number of conflict years

At the most basic level, the answer, if somewhat tautological, is that states vary with respect to the factors that produce conflict. There is a wealth of quantitative empirical scholarship in international relations and comparative politics indicating the importance of some country-level characteristics in producing conflict (Collier & Hoeffler 2004, Fearon & Laitin 2003, Hegre & Sambanis 2006, Sambanis 2004). Low levels of economic development, weak governance, social and economic inequalities, ethnic divisions, large populations, mountainous terrain, natural resources, among others, have been argued to increase the risk of civil conflict. As such, states with the greatest abundance of these characteristics should experi-

ence the most conflict. However, we observe substantial variation in the regularity of civil conflict that cannot be well explained by these initial conditions alone. Why, for example, has Kenya never experienced a civil war, while neighboring Ethiopia and Somalia have had repeated and lengthy episodes of conflict, despite similar structural risk factors?

These results have led many to conclude that civil war *itself* makes successive episodes of conflict more likely. As Collier *et al.* (2003) suggest “once rebellion has started it appears to develop a momentum of its own.” While not the first to recognize the apparent cyclical pattern of civil war, the work of Collier and his colleagues has produced a well-known theoretical explanation for this recursive process, namely, the conflict trap. Specifically, Collier *et al.* (2003) identify several ways in which civil war makes conflict more likely: (i) reversing development, (ii) triggering emigration and diasporas, (iii) leaving a persistent and damaging military lobby, (iv) changing the balance of interests and intensifying hatreds. Each of which has been shown elsewhere to increase the risk of civil conflict (Blattman & Miguel 2010). In effect, fighting a civil war generates or exacerbates conditions which make civil war more likely, as the consequences and causes of civil war are often the same. Thus, once on a path of conflict, it can difficult for states to achieve the conditions necessary to promote sustained peace.

Earlier work exploring the conflict trap focused on the ways in which aspects of the prior conflict contribute to the likelihood of future war. Walters (2004) summarizes these studies as those which have focused on: i) why the original war began, ii) how the war was fought, iii) how the war ended. More recently, a growing literature has examined the ways in which conflict promotes recurrent conflict via the mechanisms identified above (e.g., economic destruction, societal polarization and militarization, etc...), or what Hegre *et al.* (2011) calls the production of “conflict capital” (Collier *et al.* 2003, Hegre *et al.* 2011, Walters 2004). In particular, the majority of this work focuses on the ways in which conflict destroys infrastructure, reduces investment, diverts capital from productive sectors of the economy, and can drive away intellectual capital. Each of these reduces the economic capacity of the state following the war, which itself is widely considered one of the most robust determinants of civil conflict (Hegre & Sambanis 2006). Collier *et al.* (2003) argue that, in addition to these

observable consequences, many unobservable factors also emerge from fighting: increased animosity and hatreds, the production of a solidier class, the destruction of social capital. Each is argued to increase the likelihood the we observe repeated conflicts within the same country.

However, I argue an additional conflict-generating consequence of war has gone underanalyzed in the current literature on conflict inertia: contagion.⁴ While the literature conflict persistence has focused on the ways in which fighting creates conditions favorable to additional conflicts in *that* state, the literature on contagion identifies the ways in which civil war increases the risk of additional conflicts in *neighboring* states (e.g., Diehl 1991, Gleditsch 2002, Lake & Rothchild 1998, Most & Starr 1980, Starr & Most 1983, Ward & Gleditsch 2002). Civil wars trigger refugee flows to surrounding countries, depress regional economic capacity and investment, and introduce combatants, weaponry, and ideologies to the region, each of which is argued to make conflict more likely in neighboring countries (Buhaug & Gleditsch 2008, Murdoch & Sandler 2002, Salehyan & Gleditsch 2006). In sum, civil war increases the likelihood of conflict regionally.

Thus, the current literature has suggested two independent processes whereby civil conflict causes additional conflict: persistence and contagion. Deviating from this work, I argue that these are part of a single system of positive feedback, with persistence augmenting contagion and contagion augmenting persistence. First, protracted or repeated conflicts should make regional spillovers more likely. Longer or recurrent conflicts simply provide more opportunities for instability to spread. Each of the mechanisms through which conflict is argued to spread – e.g., refugee flows, economic disruption, weapons proliferation – is more likely to be triggered in environments with persistent conflicts.⁵ Second, the contagion of conflict regionally increases the risk of conflict persistence, as each state’s risk of conflict influences, and is influenced by, neighboring states. That is, the likelihood of conflict in any single state is a function of both its independent risk of conflict (e.g., country-level factors,

⁴Where noted it is given a cursory treatment – (Collier *et al.* 2003), for example, devote less than a page of a more than 200 page manuscript to the subject – and no work, to my knowledge, has linked contagion to conflict persistence itself.

⁵This argument, if unstated, is implicitly made by much of the civil war contagion literature already when they use a spatial lag of conflict incidence in models predicting civil war onset.

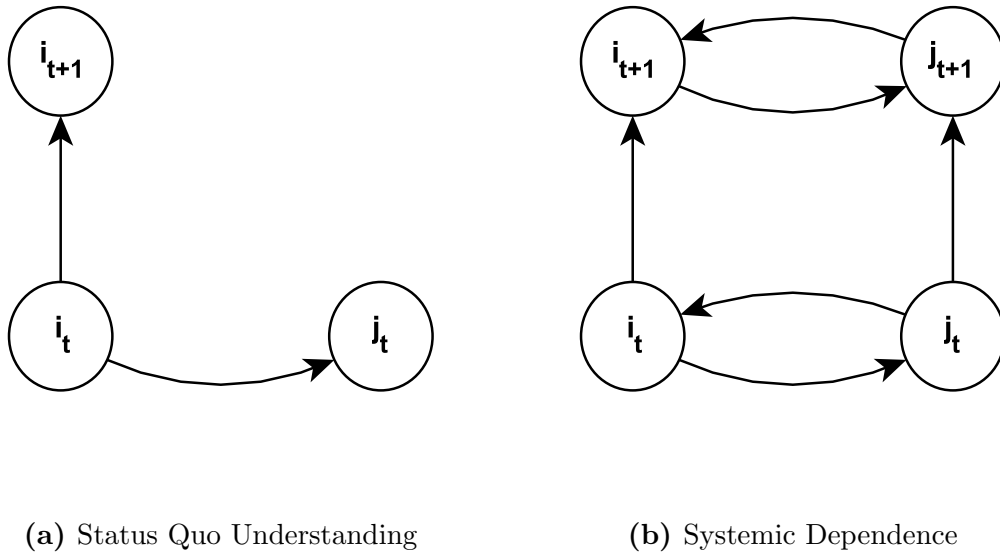
conflict history, etc...) *and* the risk of conflict in their neighbors. Thus, *ceteris paribus* states surrounded by at-risk countries are at a greater risk of (re-)experiencing conflict themselves.

To further clarify how contagion serves as a multiplier exacerbating persistence, consider the impact of an exogenous increase (x) to the risk of conflict in country i at time t . In [Figure 3.2](#), I depict expected impact of a such a shock given contemporary understandings of conflict dynamics [Figure 3.2a](#), and my argument on systemic dependence [Figure 3.2b](#). With our current understanding, this shock increases the latent propensity (i.e., risk) of civil conflict in i_t which then increases the risk of conflict in i_{t+1} directly via temporal persistence. However, I argue that such a theory offers an incomplete account of the conflict dynamics and, as a consequence, underestimates the propensity for conflict in i_{t+1} – that is, the persistence of conflict – in at least 2 ways. First, it neglects the simultaneous spatial feedback through the increase to the regional risk of conflict, with any increase x having a direct *and* indirect (i.e., spatial) effect on i 's conflict propensity.⁶ That is, an increase to i 's risk (direct) increases j 's risk which then feeds back into i 's risk (indirect), these are known as second-order (or ‘echo’) effects.⁷ Secondly, the increase in j 's risk at t increases the propensity for conflict in j_{t+1} (via persistence), which itself increases the risk of conflict in i_{t+1} (via contagion), and so on and so forth. In sum, we underestimate the propensity for conflict in i_{t+1} by neglecting the way in which the persistence of conflict is augmented through this positive and recursive spatial feedback loop.

⁶Conventional conflict contagion theories also neglect this dynamic.

⁷In short, the intuition here is that any state is itself a neighbor of its neighbor. Therefore, in the same way effect of a change diffuse to non-contiguous states they also feedback to the originating country.

Figure 3.2: The Dependence of Conflict in Space and Time



This system of feedback can produce conflict (or peace) prone regions, as any initial increase to the risk of conflict (or the observation of civil war in any state) can increase the risk of conflict amongst a set of interconnected states many times over. This suggests that the so-called ‘conflict-trap’ may be better understood as a *regional* phenomenon, with each state’s individual probability of conflict multiplied through those of their neighbors.

There is strong anecdotal evidence for the importance of these dynamics in the production of conflict prone *regions* in Sub-Saharan Africa and Southeast Asia. Several countries in the Great Lakes region of (East) Africa (e.g., DRC, Uganda, Rwanda, Burundi, Tanzania) have seen persistent and repeated civil conflicts that appear related. [O’Loughlin & Raleigh \(2008\)](#) argue that this set of conflicts “highlight how many current civil wars are not state-specific, but related and supported by a host of external conditions” (10). Ciaran Donnelly, head of International Rescue Committee effort in Uganda, reiterates “in general, what you can see in this whole region is a set of interrelated conflicts” ([McHugh 2005](#), 5). Furthermore, the West Africa region (Sierra Leone, Liberia, Guinea, Senegal) seems to exhibit similar regional conflict dynamics. In 2003, *The Economist* reported: “West Africa’s civil wars are usually reported as tragedies befalling individual states...In fact, all these wars

are intertwined, and it is impossible to understand one without reference to the others.” Thus, countries located in such clusters have persistently higher risks of conflict than would be expected from any single country.

Though these accounts suggest the importance of spatial interdependence in producing persistent regional conflict, little empirical work in political science has systemically explored this question. Therefore, in following sections, I test the extent to which spatial dependence contributes to recurrent civil conflict.

3.2 RESEARCH DESIGN

To test the preceding propositions, I estimate a series of temporally and spatially lagged probit models of civil conflict using maximum simulated-likelihood with recursive-importance-sampling (RIS).⁸ As the method has already been discussed at length in [chapter 2](#), I limit any explicit discussion of the setup, formula, and estimation, focusing instead on describing the data in the model and briefly discussing the leverage MSL-by-RIS offers me in answering my theoretical question.

The general form of the time-series cross-sectional model is given in [Equation 2.9](#). Here, y_{it} is *Civil Conflict*, which takes the value of 1 for any country-year in which a civil conflict occurs and 0 otherwise, evaluated for all countries in Sub-Saharan Africa from 1961 to 2008.⁹ The data on civil conflict incidence comes from the UCDP/PRIO Armed Conflict Database ([Harbom & Wallensteen 2007](#)), which defines civil conflicts as violent incidents between a state government and organized opposition which result in at least 25 deaths. The spatiotemporal multiplier, $(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}$, captures the extent to which there is

⁸Note that the computational costs of estimating these models is severe, with each doubling of N decreasing the the rate of convergence by about 3.5 times. As such, a model with 1,800 units – e.g. therefore, a 1,800-dimension integral to be simulated) – can take a full day to estimate.

⁹There are three reasons I limit the current analysis to Sub-Saharan Africa. First, substantively it is the most important region for understanding the dynamics of civil conflict. Second, [Achen \(2002\)](#) suggests we should select samples which enable us to reduce the need for extraneous independent variables. Finally, the computational costs for estimating models with a global sample are immense, with each model taking days to compute.

spatial and temporal dependence in civil conflict incidents. In this analysis \mathbf{W} is a row-standardized binary-contiguity matrix, with each element of w_{ij} identifying whether states i and j share a common border as defined using the Correlates of War Direct Contiguity v3.1 data (Stinnett *et al.* 2002). Therefore, the spatial lag ($\mathbf{W}\mathbf{y}^*$) captures the effect (ρ) that civil conflicts in neighboring states (Σ_j) have on the likelihood of a civil conflict in country i . Similarly, the time lag ($\mathbf{L}\mathbf{y}^*$) – with \mathbf{L} coded as the standard single-period lag discussed in chapter 2 – captures the effect (ϕ) of the previous periods likelihood of conflict on its current value, within unit. Positive and significant estimates of these parameters (ϕ and ρ) would offer support for the dependence of civil conflict across time and space respectively, with significant effects in both indicating a *regional* conflict trap.

Using MSL-by-RIS allows us to model the dependence of binary outcomes in terms of latent- y^* , which has three principle advantages. First, it allows us to directly model the auto-regressive process, analogous to familiar time series models of continuous outcomes.¹⁰ That is, we can model temporal auto-dependence, the dependence of a state’s current conflict propensity on its prior conflict propensity. Importantly, this provides us a flexible strategy to model conflict dynamics, including the effect of unobserved factors on the realization (and recurrence) of conflict.¹¹ Despite these benefits, computational difficulties have prevented the wider utilization of such models. Instead, civil war scholars have largely evaded the issue by using event-history approaches suggested by Beck *et al.* (1998) and Carter & Signorino (2010).¹² As discussed in chapter 2, while these evade the bias induced by erroneously assuming independence, they do not directly model the temporal-autoregressive process. Furthermore, they cannot effectively incorporate conflict dependence across space which can only consistently be modeled via latent- y^* .

Second, although spatial models of civil conflict have become increasingly frequent, these all involve the inclusion of a spatial-lag of neighboring conflict as an exogenous right-

¹⁰Here I model these dynamics as an AR(1) process, but this strategy can accommodate a range of AMRA model.

¹¹While Jackman (2000) suggests this is an ideal approach to capture these effects, directly modeling unit effects (in addition to possible dependence in the outcome) seems a more complete approach. This is a possibility I currently explore in co-authored research elsewhere (Cook, Hays, and Franzese 2013).

¹²The notable exceptions being the working papers of Jackman (2000) and Beck *et al.* (2001) discussed in chapter 2

hand side regressor in a standard logit or probit model (e.g., Buhaug and Gleditsch 2008; Saleyhan and Gleditsch 2006). Such analyses are *all* necessarily biased, having failed to account for the endogeneity induced by including the spatial lag simultaneously.¹³ That is, in the same analysis these models regress y_{jt} on y_{it} and y_{it} on y_{jt} , thereby producing results that are badly biased. Time lagging the spatial-lag, as is occasionally done, only evades this bias if we assume there is no within-year (i.e., simultaneous) effect of neighboring conflict, which history suggests is rarely, if ever, the case. Failing this, we can only model spatial effects consistently through the latent variables or errors (Heckman 1978). MSL-by-RIS allows me to achieve this, thereby providing the *first* consistent and efficient estimates of spatial effect of civil conflict in the literature.¹⁴

Finally, modeling both temporal and spatial dependence in latent- y^* provides the most effective means of discriminating between their individual effects and exploring their joint impact. Notably, it is the only estimation strategy which provides consistent estimates of both dependence parameters while employing a consistent theoretical logic.¹⁵ Furthermore, it easily allows us to decompose the effect of the regressors into short- and long-run, spatial and non-spatial, thereby providing clearer insight into the means by which they influence conflict.

In addition to modeling the dependence in civil conflict, I include a battery of additional covariates (\mathbf{X}) to account for several of the well-known country-level determinants of civil conflict. *GDP* (\ln) is the natural log of per capita GDP, which has been argued to influence state capacity to prevent rebellion, with greater levels of development reducing the risk of conflict. *GDP growth* is the percentage over per capita GDP growth over the previous year, with higher rates of growth argued to make conflict less likely by increasing the opportunity

¹³Franzese *et al.* (2014) provides evidence of the extent of this bias using monte carlo analysis.

¹⁴Somewhat more technically, with interdependent outcomes the probabilities for each unit are necessarily related, meaning that instead of maximizing the log of the product of N marginal distributions, we face a single N -dimensional integral. High-dimension integrals lack analytical solutions, instead a numeric or simulation strategy is required. My approach builds on a sampling strategy discussed in Train (2009) and outlined specifically in Franzese *et al.* (2014).

¹⁵For example, one could construct a consistent hybrid model using a regime-switching model (for time) and MSL-by-RIS (for space). While econometrically feasible, there is a clear incongruence in the underlying theory of dependence in this model (using observed-outcomes for time by latent-propensities for space). Furthermore, such an approach would alleviate none of the computational challenges confronted in estimating these models (principally one of the chief reasons for pursuing alternative approaches).

costs of joining rebellion. *Pop* (\ln) is the natural log of national population, which has been argued to increase the likelihood of conflict, making recruitment easier and deterrence strategies more complicated. All three of these measures come from the Penn World Table 8.0 data (Heston *et al.* 2012). Lastly, I include a binary measure of regime type, *Democracy*, from Cheibub *et al.* (2010) Democracy-Dictatorship data, which should reduce the risk of conflict by providing non-violent means with which to pursue political grievances. For each of these covariates, I estimate both the short- and long-run effects, which, to my knowledge, have not been calculated in any of the civil conflict literature to date.¹⁶

3.3 RESULTS

The results from my model suggest considerable support for the dependence of civil conflict across time and space (Table 3.1). Model 1 provides the estimates of the regressors when we assume conflict incidence is independent, with each of the regressors achieving traditional levels of significance in the expected directions. However, as expected, there is a persistent effect of civil conflict, as the temporal lag is highly significant (Model 2). Though we cannot yet speak to the substantive significance, an issue I turn to shortly, a coefficient of 0.710 suggests considerable dependence in the model. Furthermore, this model provides insight into how the regressors influence civil conflict. Notably, the parameter estimates for *GDP* and *Democracy* both fail to obtain traditional levels of significance when lagged- y^* is included in the model, suggesting that the impact of these variables is predominantly through their long-run effects.¹⁷ As in continuous-outcome models, the total effect is now given by $\frac{\beta}{1-\phi}$, which is -0.135 and -0.324 for *GDP* and *Democracy* respectively, which is quite similar to the estimated values given in the independent model.

¹⁶More completely, the inclusion of latent- y^* allows me to estimate the short- and long-run, spatial and non-spatial effects.

¹⁷As in OLS models with lagged-dependent variables the coefficient estimates for the variables indicate their short-run or immediate effect. See Keele and Kelly (2006) for an thorough discussion of these issues with continuous outcomes.

Table 3.1: Dependence of Civil Conflict in Sub-Saharan Africa

	Model 1 Independent	Model 2 Temporal	Model 3 Spatial	Model 4 Spatiotemporal
<i>Time Lag</i>	-	0.710*** (0.023)	-	0.702*** (0.023)
<i>Spatial Lag</i>	-	-	0.129*** (0.042)	0.061** (0.028)
<i>GDP (ln)</i>	-0.143*** (0.051)	-0.039 (0.036)	-0.131** (0.051)	-0.034 (0.036)
<i>Pop (ln)</i>	0.383*** (0.032)	0.142*** (0.024)	0.381*** (0.032)	0.147*** (0.024)
<i>GDP Growth (%)</i>	-1.061** (0.413)	-1.026** (0.410)	-1.090** (0.414)	-1.079** (0.414)
<i>Democracy</i>	-0.353*** (0.119)	-0.094 (0.090)	-0.363*** (0.119)	-0.097 (0.090)
<i>Constant</i>	-3.218*** (0.467)	-1.318*** (0.349)	-3.174*** (0.469)	-1.332*** (0.352)
<i>N (states)</i>	1780(41)	1780(41)	1780(41)	1780(41)

SEs in parentheses. ***significant at 1%; **significant at 5%; *significant at 10%.

There is also strong evidence for a spatial and spatiotemporal dependence in the model. In Model 3, I include the spatial lag, but omit the temporal lag. While, of course, not as strong the temporal dependence, the effect of spatial dependence is still positive and significant. Though there are less noticeable changes to the included regressors, our interpretation of these estimates again changes following the inclusion of a dependence parameter. In this case, the coefficients represent the direct (i.e., non-spatial) first-period effect of the regressors.¹⁸ When I include both the temporal and spatial lags in the model (Model 4), we see that both are positive and significant, indicating that there is temporal *and* spatial dependence in civil conflict. This provides support for my argument about the need to account for both of these complementary and reinforcing channels of conflict propagation.

Yet, these findings can only tell us so much. Ultimately, we are interested not only in properly estimating the parameter coefficients, but in calculating substantive effects indicating the increased risk conflict given previous or neighboring civil conflict. Despite this,

¹⁸Total effects can be calculated using the direct coefficient and the spatial filter $(\mathbf{I} - \rho\mathbf{W})^{-1}$.

few studies have explicitly calculated the substantive effects of conflict persistence and even fewer for contagion.¹⁹ As discussed in [chapter 2](#), estimating these quantities with a spatial or spatiotemporal probit model is not straight-forward, as it is complicated by the same nonlinearities in and interdependence across outcomes which plagued estimation. Therefore, I use the simulation strategy detailed in [chapter 2](#) to estimate conditional counter-factual effects. To review, I sample from the distribution of disturbances (ϵ) using the reduced-form model and generate simulated conflict episodes according to the measurement equation:

$$\Pr(y_{it} = 1) = \Pr(-(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}\mathbf{X}_{it}\boldsymbol{\beta} < (\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}\epsilon_{it}) \quad (3.1)$$

That is, the outcome (y_{it}) is a function of whether the reduced-form disturbance, $(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}\epsilon_{it}$, is greater (or less than) the negative of its reduced-form cutpoint, $(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}\mathbf{X}_{it}\boldsymbol{\beta}$. From this new set of y 's, we can estimate the counter-factual probabilities of interest and then estimate frequencies over outcomes. For example, to calculate the contemporaneous spatial effect we would compare:

$$\Pr[y_{it} = 1|X, \mathbf{W}, \mathbf{L}, y_{jt} = 1] \text{ and } \Pr[y_{it} = 1|X, \mathbf{W}, \mathbf{L}, y_{jt} = 0] \quad (3.2)$$

Simply put, we can estimate the difference in the probability that y_{it} experiences civil conflict conditional on the outcome in y_{jt} .

[Figure 2.2](#) helps to clarify this point further. In short, the reduced-form cutpoints, $(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}\mathbf{X}\boldsymbol{\beta}$, for two selected units, divide the plane into four possible quadrants. For both units, we take a draw from the reduced-form disturbance, $(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}\epsilon$, which jointly identify an x-y coordinate (a point)²⁰. The relation of this point to the cutpoints (i.e., the quadrant it is located in), indicates one of four possible outcome profiles. For example, if the point is in quadrant I, then both i and j are above their respective reduced-form cutpoints and $y_{it} = 1$ and $y_{jt} = 1$. After repeating this many times, we can calculate conditional

¹⁹A notable exception is the working paper by [Hegre et al. \(2011\)](#) which is notable in two respects: i) it uses a transition (i.e., regime-switching) model to capture the persistence in conflict and ii) it provides country-specific effects estimates using predicted probabilities. However, in addition to other differences, they treat the spatial lag of conflict as exogenous, resulting in biased estimates of both the spatial and temporal parameters for the reasons outlined above in the text.

²⁰To clarify further, the disturbances are drawn for the entire vector of units in the model, but for this illustration I simply utilize the slice corresponding to the two units of interests

relative frequencies from the ratios of the quadrant counts which provide estimates for the conditional probabilities. Specifically, the probabilities given in (3.1) are estimated by:

$$\begin{aligned}\Pr[y_{it} = 1|X, \mathbf{W}, \mathbf{L}, y_{jt} = 1] &= \frac{\text{Quadrant I Count}}{\text{Quadrant I Count} + \text{Quadrant IV Count}} \\ \Pr[y_{it} = 1|X, \mathbf{W}, \mathbf{L}, y_{jt} = 0] &= \frac{\text{Quadrant II Count}}{\text{Quadrant II Count} + \text{Quadrant III Count}}\end{aligned}\tag{3.3}$$

With the difference between the two estimated probabilities giving us the increased (decreased) probability that y_i is equal to 1 given the outcome in y_j (i.e., the conditional contemporaneous spatial effect). This same approach extends straightforwardly for estimating the substantive effect of conflict persistence, we simply estimate the probabilities in (6.3) conditioning on counter-factual realizations of the outcome in y_{it-1} as opposed to y_{jt} .

This strategy enables us to answer hypothetical counter-factual questions about the extent to which neighboring and prior conflict make the realization of conflict more likely for specific countries. Even a cursory review of recent political coverage uncovers the frequency with which such questions are (implicitly) posed: “UN Chief Fears Resumption of Civil War In Ivory Coast,” “New Civil War Feared in Sudan As Town Empties,” “Fears mount that Côte d’Ivoire conflict could spill to Liberia,” “Zambia Concerned About DRC conflict spillover,” “The conflict in Mali could be creating a ‘ticking time bomb’ for neighbouring Western Sahara.” Current research offers us little purchase over these questions, which are important for both academics and policy makers alike. However, these are exactly the type of conditional effects we can estimate via simulation.

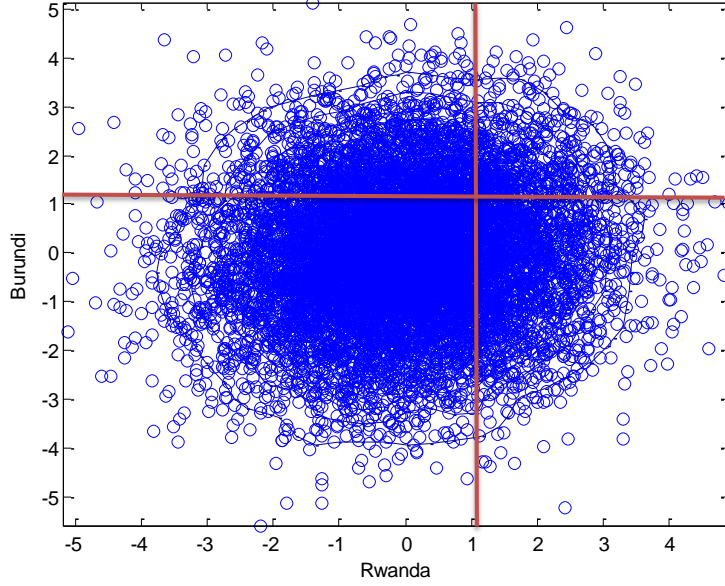
Using this strategy, I first explore the extent to which the outbreak of conflict in Rwanda in 1990 increased the risk of civil conflict in neighboring Burundi. The history between these locations runs deep, to a time since before they were states, having each emerged from German (and later Belgian) colonial control. Both achieved independence in 1962, followed shortly thereafter by repeated bouts of ethnic conflict between Hutu and Tutsi factions. Rwanda and Burundi’s close proximity and common ethnic divisions have frequently resulted in both countries, directly or indirectly, playing a role in any conflict

in the other (Chossudovsky 1996). In 1990, the – primarily Tutsi – Rwandan Patriotic Army invaded Rwanda from Uganda, triggering a series of events which culminated in the Rwandan Civil War. Given the proximity of Rwanda and Burundi, what impact did this initiation of fighting have on the likelihood of conflict in Burundi? With respect to the model, the question becomes: given that Rwanda’s reduced-form disturbance is above/below the negative of its reduced-form cutpoint, what is the probability that Burundi’s reduced-form disturbance will be above/below the negative its reduced-form cutpoint?

Using the approach indicated above, I draw 10,000 times from a $N(0,1)$ for each of the observations in the sample and then multiply them through the spatiotemporal filter to estimate the reduced-form disturbances. The reduced-form cutpoints, which are generated by multiple $X\beta$, for Burundi and Rwanda by the spatiotemporal filter, are -1.1395 and -1.0907 respectively. Therefore, a civil conflict occurs if the reduced-form disturbances are greater than 1.1395 and 1.0907. The distribution of the disturbances in relation to the (adjusted) cutpoints is presented in Figure 4. In these simulations, Burundi experiences a civil conflict 19.7% of the time when Rwanda does not (1537/7790), yet 24.6% of the time when Rwanda is also embroiled in civil conflict (544/2210). Thus, conflict in Rwanda increases the risk that Burundi will experience conflict by 4.88%. To calculate our uncertainty about these estimates – that due to sampling uncertainty about the coefficient values – we can resample the model parameters (β, ρ, ϕ) from a multivariate normal using the estimated means and variance-covariance matrix from the preceding estimation. Doing so 100 times indicates that with 95% confidence the immediate effect estimate lies between 2.27% and 5.94%

These results confirm the presence of significant and substantive positive spatial interdependence interdependence in civil conflict. However, we are interested in whether this dependence also contributes to regional conflict persistence. Therefore, I expand the analysis in two ways. First, I include an additional country, the Democratic Republic of Congo (DRC), which gives greater leverage over the influence of multiple neighboring conflicts, as is frequently the case in such clusters. Second, and more importantly, I expand the analysis in time to look at how spatial and temporal dependence jointly contribute to the persistence

Figure 3.3: Contagion: Contemp. Spatial Effects of Conflict(Burundi & Rwanda)



of conflict.

The strategy for multiple countries is largely the same as in (6.3), except now we must expand to three, or more, dimensions. Now the reduced-form disturbances for the three units, $(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}\epsilon$, identifies (X, Y, Z) locations on the Cartesian coordinate system. As before, the location of the point relative to the reduced-form cutpoints, now indicated by planes, specifies an outcome profile, with $\{1,1,1\}$ indicating conflict in all 3 states. Using this strategy, we can calculate multiple conditional frequencies. For example, if we just wanted to compare the case in which both neighbors experience conflict to that in which neither do, we calculate:

$$\Pr[y_{it} = 1 | X, \mathbf{W}, y_{j1t} = 1, y_{j2t} = 1] = \frac{\text{freq}\{1,1,1\}}{\text{freq}\{1,1,1\} + \text{freq}\{0,1,1\}} \quad (3.4)$$

$$\Pr[y_{it} = 1 | X, \mathbf{W}, y_{j1t} = 0, y_{j2t} = 0] = \frac{\text{freq}\{1,0,0\}}{\text{freq}\{1,0,0\} + \text{freq}\{0,0,0\}}$$

Though, of course, more elaborate sub-comparisons are possible – the effect of conflict in one and not the other – through different calculations of the conditional probabilities and subsequent frequencies.

Figure 3.4: Contagion: Regional Spatial Effects of Conflict(Rwanda, Burundi, & DRC)

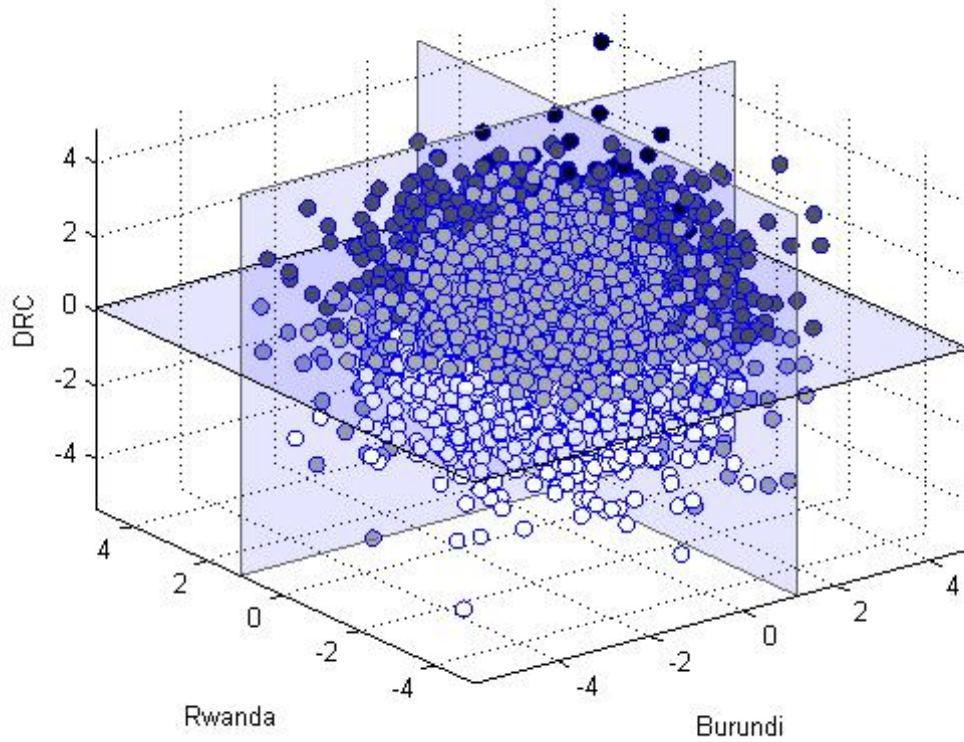
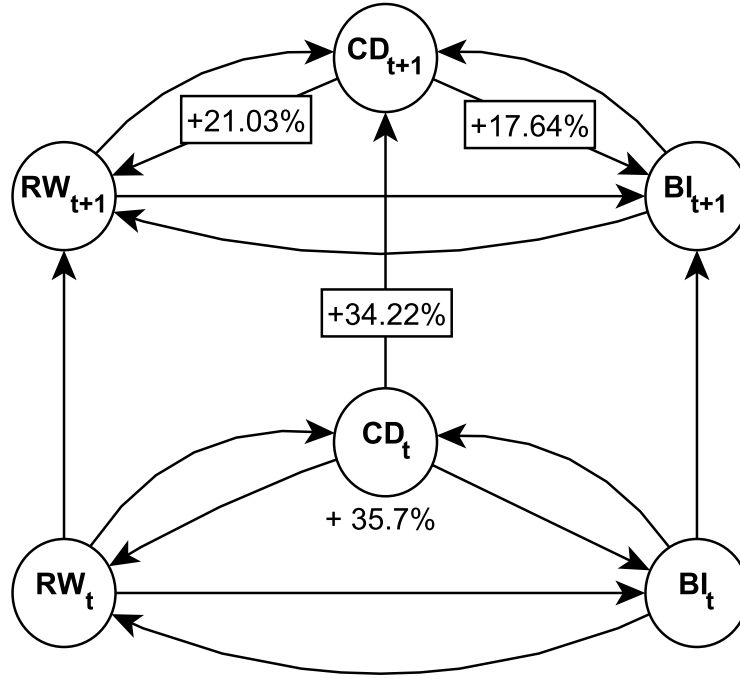


Figure 5 shows the results of this strategy for Rwanda, Burundi, and the DRC, with the planes representing the cutpoints for each of the countries and the points representing the locations of the draws of the disturbances. To facilitate comparison, I note the location of the points in relation to the cutpoints by their color, with white indicating the point is beneath all of the cutpoints (e.g, no conflict, located in the bottom front-most region) and the darkest points exceeding all 3 of the cutpoints (i.e., conflict in all, located in the top-rear region). Calculating the effects indicated in (3.4) for the DRC is achieved by calculating the frequency of points above/below the horizontal plane in the foremost region, and comparing those to the frequency of points above/below that plane in the rear region. The results from this indicate that the probability of conflict in the Democratic Republic of Congo increases

by 35.7% when both Rwanda and Burundi are currently undergoing conflict (as compared to when neither are).

Figure 3.5: Conflict Trap: Persistence & Contagion (Rwanda, Burundi, & DRC)



These effects are also represented in the complete spatiotemporal regional feedback cycle, presented in Figure 6. In this analysis, I also explore the impact that such a transition would have on the risk of conflict in the Democratic Republic of Congo (CD in the illustration) moving forward. Conditional on conflict occurring at time t – made 35.7% more likely due to spatial feedback – the risk of conflict in the DRC in the next period is 34.22% greater. To see how these effects feedback yet again, I calculate how this increased risk contributes to the risk of conflict in Rwanda and Burundi in $t+1$. Conflict in the Democratic Republic of Congo (at $t+1$) increases the risk of conflict by 21.03% in Rwanda and 17.64% in Burundi. Which, if it results in conflict in either, feeds back again into an increased risk of conflict in the DRC. This illustrates the means by which several interconnected countries can experi-

ence persistence which arises from contagion, as any risk of conflict is further heightened via these simultaneous spatial channels.

3.4 DISCUSSION

“Once rebellion has started it appears to develop a momentum of its own,” moving through both time and space (Collier *et al.* 2003). Evidence presented here suggests that the persistence of conflict occurs not just because of its impacts on country currently fighting, but also because of the consequences that fighting has on its neighbors. I have shown here that such conflict increases the risk of conflict in neighboring countries which, in turn, increases the risk of conflict in the originating state. Furthermore, this increases the probability of conflict in neighboring states during future periods, which again spatially feeds back into a heightened likelihood of persistence or recurrence in the original conflict country. This system of simultaneous and sequential positive feedback explains the clusters of states with recurrent conflict (i.e., conflict regions) that we frequently observe. In sum, the evidence supports my argument for *regional* conflict traps, wherein states become locked in persistent and interrelated civil conflicts. This not only offers a more complete understanding of conflict dynamics, but helps to explain why peace has been so difficult to achieve in some areas.

Furthermore, these results indicate that once spatiotemporal dependence is modeled, traditionally important causes of conflict (Democracy and GDP) are no longer significant. The finding for GDP per capita is particularly interesting, as this is widely considered one of the most robust relationships in civil war (Hegre & Sambanis 2006). From my findings it appears instead that GDP emerges as a significant predictor of conflict *because* of a failure to adequately capture the temporal and spatial dependence. Why should this matter? First, countries that are poor also experience more (and repeated) conflict, meaning treating these events as serially independent artificially inflates the relationship between GDP and conflict. Second, poor states are spatially distributed in a pattern similar to conflict prone states – recall the African states discussed in the last section – meaning, again, that failing to capture

this spatial dependence in the outcomes will erroneously inflate the relationship between GDP and conflict. Given the significance assigned to development as an instrument for promoting peace, this is a very important relationship to properly understand. Therefore, in [chapter 6](#), I explore an additional explanation, that unobserved constant unit factors make some states more likely to experience both low development and conflict, suggesting the relationship between the two is spurious. Ultimately, it will be important to distinguish completely between these three possible explanations for the “developmental peace” – is it low development, conflict dependence, or unobserved unit heterogeneity which causes some states to experience conflict so regularly? – but parsing out these individual effects is not easily achieved with linear models, let alone in non-linear ones [Nickell \(1981\)](#). As such, in [chapter 7](#), I propose dynamic panel models for binary outcomes which may allow us to discriminate between these accounts.

In addition to that work, several other questions need to be answered in the future to better understand how and why states become trapped in conflict. First, we need to better understand the role of exogenous-external factors in triggering and continuing these cycles of conflict. In future work, I plan to explore the affect of commodity price shocks on conflict resumption. Are post-conflict countries at greater risk of experiencing conflict following an economic shock, and do such shocks, in turn, indirectly impact the likelihood of conflict in surrounding states. This is particularly important to the extent that these exogenous-external factors simultaneously impact several states in the region – e.g., clusters of states export a similar resource, neighboring states experience similar rainfall levels, etc... – as these common shocks could also explain the regional clustering of conflict episodes. Second, how does the risk of conflict contagion impact the decision for states to become directly involved in neighboring conflicts (e.g., intervention), and does such intervention actually decrease the risk of conflict contagion? Scant work in the current literature on intervention has explored these questions (notably [Kathman 2010; 2011](#)), and none has accounted for the endogenous dynamic evident in this relationship. Estimating models to fully capture this dynamic is challenging, however, for the reasons discussed in [chapter 2](#).

Does this analysis suggest that countries in ‘bad neighborhoods’ are fated to continuously suffer the ills of war? Or that states should attempt to wall themselves off from their neighbors? Far from it, the same positive feedback cycles actually provide a possible means out of the trap. Just as cluster of conflict-prone states can emerge, so to can a pocket of peace-prone states (e.g., West Europe). That is, as any state begins to reduce its risk of conflict, this also feeds back and reduces the risk of their neighbors conflict, and so on and so forth. What is needed are strategies for reducing the risk of conflict such that those gains might also diffuse. Obviously, such research is already underway, with political scientists and policy makers both exploring strategies for building social capital, improving economic investment, promoting sound governance. My analysis here suggests that while states currently in clusters of conflict have it the most difficult, any successes achieved there may also be felt many times over. Furthermore, it also indicates that the costs of failing to do so are greater than previously assumed, with several states affected, which should serve to bolster our desire to root out conflict.

4.0 MODEL MISSPECIFICATION: POLITICS AND THE CONTAGION OF FINANCIAL CRISES

Finally, you have broader considerations that might follow what you would call the “falling domino” principle. You have a row of dominoes set up, you knock over the first one, and what will happen to the last one is the certainty that it will go over very quickly.

— Dwight D. Eisenhower, 1954

While, civil conflict may require borders to spread, this is not the case for many other phenomena. The relationships between states are complex and multidimensional, with states varying in their respective levels of economic, political, and military integration with one another. Each creating a tie with the potential to cause events in one state to affect events in another. While early work in spatial statistics – focusing on the proximity of agricultural plots – did not often require complex measures of spatial association, spatial econometric analysis is now applied to estimate a variety of models which include more complex spatial structures. In International Relations states are connected through a multiplicity of overlapping direct (e.g., alliances, trade, IOs, treaties) and indirect (e.g., common political institutions, language, colonial origin) ties. As such, conventional spatial- and spatiotemporal autoregressive models (i.e., single spatial-lag) are biased, both *over*-estimating the effect of the included spatial-lag and *under*-estimating the total spatial dependence in the data.

One research area where this has proved problematic is the study of panics and manias. Over the last several decades, financial crises have become an increasingly common and

debilitating phenomenon.¹ Notably, there have been a rash of severe financial crises in recent years (e.g. Mexico 1994; East Asia 1998; Russia 1998; Argentina 2001), peaking with the global financial crisis of 2007-2008 (e.g. the "Great Recession") and its aftermath (e.g. Greece 2012). The severity and scope of these events has triggered a renewed interest in understanding the causes of and responses to such crises (see [Macias *et al.* 2010](#)). Though in many respects "this time is [no] different" from previous episodes, the *spread* of recent crises appears to be more common which can, in turn, make recovery more difficult ([Reinhart & Rogoff 2011](#)). As such, it is important to understand the specific means by which financial crises spread internationally. In particular, what factors increase the risk of crisis contagion?²

There is an extensive literature on the contagion of financial crises (useful reviews include [Dungey & Tambakis 2005](#), [Karolyi 2003](#), [Upper 1996](#)). To date, most of the empirical work in this literature stream has focused on the impact of extant economic ties - such as trade (among others, [Glick & Rose 1999](#)) or financial linkages ([Allen & Gale 2000](#), [Furfine 2003](#)) from one country to another. Taken together, this work indicates that, while important, direct economic ties can only offer us so much purchase on the issue of financial contagion. Consequently, many now suggest that changes to investor beliefs may spread crises above and beyond what would be anticipated from such spillovers ([Pericoli & Sbracia 2003](#)). In short, investors may respond to an initial crises by further withdrawing from the global financial system which, in turn, spreads the crisis across otherwise unconnected economies.

However, the pattern of these withdraws is neither random nor irrational ([Kindleberger 1978](#)). Instead, a financial crisis serves as a signal to investors about possible underlying structural fragility in both that country, and importantly, others ([King & Wadwani 1990](#)). From these updated beliefs, investors reduce investment in states in similarly 'fragile' economic and, I argued, *political* positions. Specifically, I argue that a financial crises in a country increases the uncertainty over the ability of states with similar political fundamentals to prevent economic crises, precipitating withdraws from politically similar states and

¹[Bordo *et al.* \(2001\)](#) find that the rate at which such crises have occurred since 1970 is matched only be the frequency witnessed during the Interwar Years (1919-39), culminating in the Greater Depression.

²[Pericoli & Sbracia \(2003\)](#) define contagion as "a significant increase in the probability of a crisis in one country, conditional on a crisis occurring in another country" (574).

triggering ‘self-fulfilling’ crises.³ This contention builds on the well-established political economy literature on the importance of the political environment for investor decision-making (Feng 2001, Haggard 2000, Jensen 2003; 2006, Li & Resnik 2003, etc...) and crisis management (MacIntyre 2001). However, rather than simply treat investor beliefs about institutions as fixed, I argue that they are consistently updated in response to recent events (e.g. conditional upon the current crisis). Moreover, the impact of this new information on beliefs may be substantial given that international investment is a notoriously low-information environment (Calvo & Mendoza 2000). As such, an event as dramatic as a financial crisis is likely to induce significant updating of investor beliefs about the risk of these political environments.

To test this proposition, I examine the effect of ‘political proximity’ - the similarity of political institutions between two states - on the contagion of financial crises using multiparametric spatiotemporal autoregressive (m-STAR) probit. While the details of this estimator are discussed in chapter 2, it has several advantages from a theoretical perspective. First, it allows us to both account for the simultaneity of crises and explicitly model several possible transmission channels (e.g., economic ties, political ties, etc...) between states. This improves over existing work, which has either failed to account for the endogeneity bias which results from such simultaneity, examined any possible transmission channels independently, or both.⁴ Second, associational or characteristic-based spatial lags (such as political proximity) offer a means of capturing whether states with similar institutions experience the onset/absence of crises in unison - i.e., whether the contagion occurs within groups - thereby allowing for a possible indirect role of domestic institutions on the realization of crises. In all, it allows us to identify and test a wider range of possible determinants of financial crises contagion.

Understanding the means by which financial crises are likely to spread is crucial to both the literature on political economy and resultant policy-making. In their review of the literature on financial crises, Allen *et al.* (2009) argue that “we need to gain a better

³The concept of political fundamentals has been discussed elsewhere in the literature (see Hays *et al.* 2003). In general, it refers to any political relevant political informational. In my subsequent empirical analysis, I focus on political institutions (formal or informal) which may influence market outcomes.

⁴To my knowledge, only De Gregorio & Valdes (2001) has included more than trade and finance linkages in a single analysis. Though, in this case, the failure to account for the endogeneity of the outcomes biases the findings and, therefore, gives us limited insight into the competing contagion channels.

understanding of the market failures that lead to financial crises...perhaps the most important of [which] is contagion” (27-28). They continue that “a full understanding of contagion is necessary before adequate policy responses can be designed.” This chapter aims to further this effort in two ways. First, it highlights the importance of political fundamentals in the contagion of financial crises. Second, it distinguishes between multiple competing forms of interdependence. Anticipating my findings, the analysis suggests that after accounting for the alternative channels of transmission, the presumed negative impact of trade is markedly reduced. This has substantial implications for our understanding of the potential downside risk of increased economic integration. Furthermore, I find support for the role of political fundamentals as a novel transmission mechanism in the spread of banking crises.

4.1 COMMON POLITICAL FUNDAMENTALS AND SIGNAL-EXTRACTION FAILURES

Since 1970, financial crises have occurred as a rate not witnessed since the Great Depression (Bordo *et al.* 2001). While these events are still relatively rare, 124 systematic banking crises and 208 currency crises have occurred across more than 50 countries (Laeven & Valencia 2012).⁵ Not only have such crises become more frequent, but also more severe than in previous financial periods (Calvo & Reinhart 1996). Afflicted economies have suffered deep and lasting recessions brought on by output contraction (Hutchinson & Noy 2006) and capital reversals (Joyce & Nabar 2009). Ultimately, the cumulative impact of these crises can be “several years of lost GDP” with output losses ranging between 63% and 302% (Boyd *et al.* 2005, 977). Given that this surge has occurred concurrently with increasing financial globalization, it has caused many to challenge the purported benefits of greater economic integration (Rodrik 1998, Stiglitz 2002).⁶ Namely, does globalization increase the risk of financial crises and, if so, how?

⁵See section 5.2 for an explicit operational definition of these events

⁶Including, among other factors, the rapid expansion of foreign direct investment since the 1980s and portfolio flows since the 1990s (Agenor 2003)

There is a vast literature on the systemic risk inherent to the modern international financial system. Since the [Morgenstern \(1959\)](#) analysis into the effects of stock market panics on foreign markets, hundreds (and potentially thousands) of articles have been written on the potential downside risk of economic interdependence. In general, scholars agree that a consequence of increased interdependence is that small perturbations in one part of the global financial system can more readily spillover and generate disturbances in other parts. Open financial markets can lead to asymmetries in access to capital, loss of macroeconomic stability, and pro-cyclical flows of capital ([Agenor 2003](#)). Moreover, deeply integrated economies are more directly exposed to volatility in capital movements, thereby increasing the risk of large capital outflows (e.g. liquidity runs) and, in turn, a financial crisis ([Rordrik 1998](#)). Simply put, domestic economies are increasingly put at greater risk of suffering macroeconomic losses from factors beyond their control.

As such, it appears that one of the most significant costs of increased integration is the threat of contagion, that is, the spread of financial instability from one country to another.⁷ A number of ‘regional’ crises in the 1990s and 2000s seemed to substantiate and augment these concerns, generating substantial academic and policy interest in better understanding the nature of contagion. Notably, the Exchange Rate Mechanism crisis (1992-93) negatively impacted several European countries, followed by the Mexican peso crisis (1994) which quickly spread to other countries in Latin America (See Map 1). Additionally, in July 1997 Thailand’s failure to defend the baht was quickly followed by the devaluation of numerous other Asian currencies (e.g. the Asian financial crisis) and the closing of financial institutions in Indonesia, Korea, Malaysia, and Thailand (See Map 2).⁸

⁷The appropriate definition for contagion is a question that has received considerable attention in the literature and will not be addressed at length here. See [Dungey *et al.* \(2006\)](#), [Pericoli & Sbracia \(2003\)](#) for useful reviews.

⁸To clarify, these maps do not represent any particular contagious episode, but rather any crisis which occurred within a specified period of time.

Figure 4.1: Financial Crises, 1990-1994 (Mexican Peso Crisis and ERM)

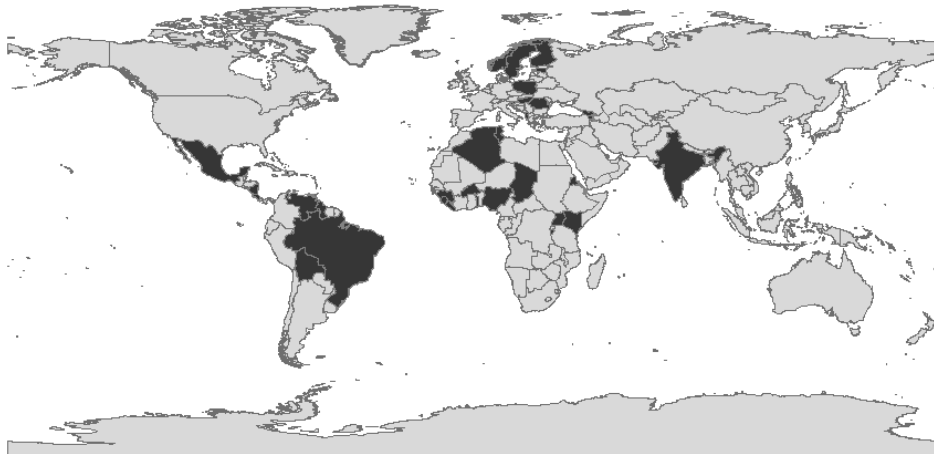
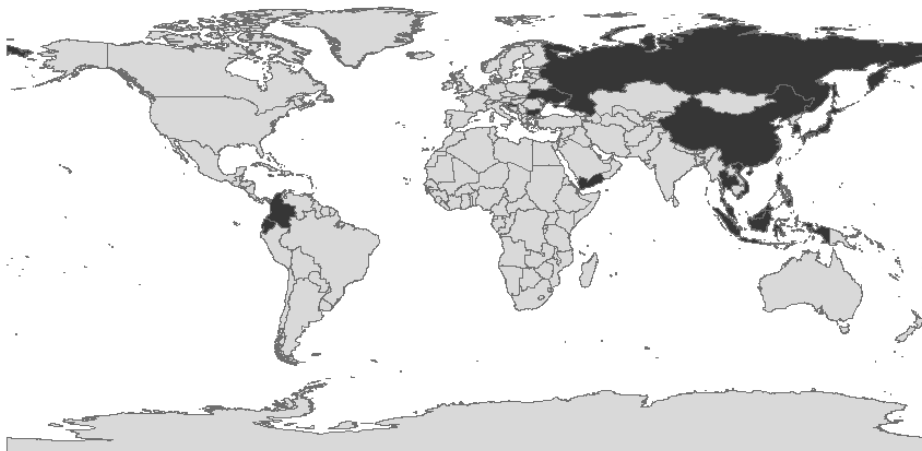


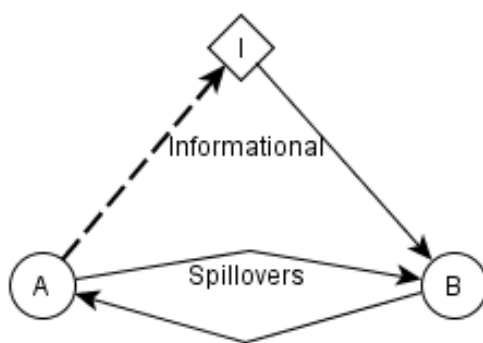
Figure 4.2: Financial Crises, 1995-1999 (Asian Financial Crises)



These events, and our failure to predict them, caused many to challenge conventional models of financial crises, which had largely highlighted the importance of sound macroeconomic fundamentals for avoiding crises ([Gorton 1988](#), [Kaminsky & Reinhart 2000](#)). In response, a large empirical literature emerged in an effort to explain the spread of these, and other, crises. This work can be split into two conceptual categories on the causes of contagion: fundamentals-based contagion and informational (or ‘pure’) contagion ([Dornbush](#)

& Claessens 2000, Dungey *et al.* 2006, Masoon 2000).⁹ Figure 4.3 presents how these two theories conceive of the spread of instability. With fundamental-based contagion, financial crises spread across direct ties between two (or more) states (i.e., ‘spillovers’), whereas with ‘informational’ contagion, some third actors witnesses an event in state A and, consequently, adopts a policy toward state B which generates instability there as well.

Figure 4.3: Competing Theories of Contagion (Spillovers vs. Informational)



Specifically, fundamentals-based contagion focuses on the cascade of crises, as dominoes, through the global financial systems via spillovers across extant real and financial ties (Calvo & Reinhart 1996, Moser 2003). First, high trade flows are argued to: *i*) make states more vulnerable to changes in market conditions in another state via changes to prices and/or the quantity trade of goods; *ii*) induce competitive devaluations; *iii*) increase the risk of speculative attacks as investors try to anticipate spillovers (Corsetti *et al.* 2000, Dornbush & Claessens 2000).¹⁰ Second, financial connections are argued to cause spillovers if, in response to some initial shock, there is a reduction in available liquidity (e.g. capital, foreign direct investment, trade insurance). Allen & Gale (2007) summarize the variety of ways in which global financial institutions are linked that may engender contagion: interbank claims, payment systems, and liquidity demand.¹¹ Finally, geographic proximity appears to

⁹Moser (2003) alternatively refers to these respectively as domino effects and information effects.

¹⁰Transmissions may also occur across real channels indirectly - between states that are not primary trading partners - via changes to export competition (Goldstein 1998).

¹¹It may be that the inability to identify a primary transmission channel - between trade and finance - is a consequence of frequently high correlation between the two (Dungey *et al.* 2006).

be a key condition in spillovers. That is, financial are argued to be a regional phenomenon, though the specific mechanism for this relationship has not been well articulated (Calvo & Reinhart 1996, Kaminsky & Schmukler 1999). While geography obviously correlates highly with trade spillovers (Glick & Rose 1999), De Gregorio & Valdes (2001) find strong regional effects which persist even after accounting for trade and economic similarities, suggesting some (potentially unobservable) neighborhood effect capturing other economic and/or political conditions which cluster by region.

In addition to direct spillovers, an alternative logic for contagion advanced in the literature – ‘pure’ or informational contagion – which centers on the behavior of international investors in response to the originating crisis. If investors believe that this crisis is likely to cause (and/or signals a greater likelihood of) additional financial crises in other markets, they may seek to preemptively withdraw investment from these states (Dornbush & Claessens 2000). These rapid capital outflows may, in turn, trigger additional financial crises. Importantly, these crises can emerge in states *unconnected* to the country of origin and, possibly, in states that would have been unlikely to experience a crisis in the absence of a change in the beliefs about the investment climate. In effect, investors construct the outcomes they had sought to avoid.

This result is largely the consequence of imperfect information held by investors and the costs of acquiring superior information (Calvo & Mendoza 2000). In general, investors lack information over: 1) fundamentals of the markets they are invested in; 2) the extent to which two economies are interdependent; 3) the cause(s) of the initial crises. Moser (2003) argues that, as a result, investors are left uncertain about the extent of the impact a financial crisis in one country has on economic outcomes in others. He summarizes two ways in which such signal extraction failures - wherein a “crisis in one country leads to an inefficient revision of fundamentals and a less accurate assessment” - may cause investors to overestimate the downside risk posed by the initial crisis on other states (163). First, investors overestimate the extent of interdependence between markets, believing spillovers to be more likely than is actually the case (Pritsker 2001). Second, investors conclude that countries similar to the crisis country possess a similar risk of undergoing a crisis. That is, there is belief among

investors that the initial crisis reveals information about the market characteristics likely to generate financial crises in other states. In this way, similar states get “lumped together” (Moser 2003).

King & Wadwani (1990) is among the first work to show evidence for such an effect in stock market co-movements, where ‘mistakes’ in asset pricing in one market are quickly transmitted to other markets. Specifically, price changes in one market are perceived as having implications for asset pricing in other markets, thereby inducing revisions across countries (Kodres & Pritsker 2002). Furthermore, Kodres & Pritsker (2002) argue that countries are at the greatest risk of contagion when information is low. Given low levels of initial information on metrics of interest, investors are likely to place undue importance on highly salient events such as the crisis. As a result, they may erroneously assume that the recent shock reveals more accurate or ‘true’ information about the stability of market fundamentals, and are therefore at greater risk of over-correcting their beliefs. Moreover, in high-cost low-information financial environments an information cascade can arise wherein investors take cues from the decisions of other actors, believing them to be better informed. Such actions prevent optimal market corrections – e.g., new investors capitalizing on the arbitrage opportunities presented – thereby resulting in Pareto-inferior outcomes such as financial crises. In particular, research has shown how perceived asymmetries of information can further exacerbate contagion, as actors update their beliefs over market fundamentals based upon the actions of other money managers and traders whom they perceive to be better informed (Calvo & Mendoza 2000, Pasquariello 2007).

To this point, however, this research has focused exclusively on economic fundamentals and been largely silent on the importance of political fundamentals. I argue we should also observe lumping over political fundamentals, thereby producing contagion between states with similar political institutions. First, it is well established that investment decisions made by market participants are determined, in part, as a function of the political environment. Investors seek economies with political fundamentals that both promote macroeconomic stability (Quinn & Woolley 2001) and limit costly uncertainty over political risk (e.g., government opportunistic behavior, expropriation, etc...). Extant literature suggests that

investment flows to states with secure property rights protections (Jensen 2003, Li 2006), strong governance (Daude & Stein 2007, Jensen 2003), government transparency (Haggard 2000), and policy stability (Feng 2001, Haggard 2000, Jensen 2006).

Second, the political economy literature has also indicated the importance of political institutions in avoiding financial crises. In particular, it is suggested that the ability to quickly respond to a shock to macroeconomic fundamentals is crucial in avoiding financial crises. MacIntyre (2001) argues that, in such instances, the heightened “uncertainty and nervousness” of investors places a premium on rapidly responding to assuage such fears and avoid market panics. Constraints on the executive may limit the ability to effectively take such actions, as leaders are required to achieve the consent of a greater number of actors with potentially divergent preferences (Haggard & McCubbins 2001, Tsebelis 1995). That is, the ability of states to take actions to promote macroeconomic stability, typically stimulus, following a shock may be either delayed or constrained as a consequence of the institutional structure, reducing the efficacy of crisis-averting policies. However, this work is contrasted by literature arguing that democracies better assuage investors because of their greater transparency (Haggard 2000), and market concerns about over-corrections, wherein states with policy flexibility may adopt unnecessary (and potentially harmful) policies because of a perceived risk of crisis. In sum, theories suggesting a direct effect of political constraints on investment seems mixed. However, there seems to be consensus that these political institutions *are* consequential in determining whether a country experiences (avoids) a crisis.

Taken together, investor beliefs over political fundamentals are important for determining investment ex ante, and for the presumed risk of those investments once a crisis is though imminent. However, research indicates that these beliefs are highly volatile. Hill (1998) argues that political risk is notoriously difficult to assess, as what constitutes a political risky environment is *highly fluid*.¹² This can result in rapid revisions of beliefs because the definition of political risk itself is potentially endogenous to the changing conditions. Therefore, Hill (1998) argues that “when assessing political risk during or shortly after a crisis, investors use the best information they have - information as to the recent crisis” (289). In

¹²Political risk is considered broadly - not simply sovereign risk - to note any variety political factors which may impact the stability, certainty, performance of the market.

effect investors mark use of (or suffer from, depending on one's perspective) the availability heuristic, wherein the probability assigned to an event is a function of the ease with which one can recall similar episodes (Tversky & Kahnemen 1973). Evidence of this bias exists for a range of financial decisions, including stock selection (Barber & Odean 2008), investment category choice (Shiller 2005), level of investment (Furfine 2003), purchasing (Folkes 1988), and analyst forecasting (Lee *et al.* 2008).

I argue that a similar dynamic exists in the behavior of international investors following a financial crisis. First, following an initial financial crisis occurrence the subjective probability of additional crises will rise to levels greater than the true risk (Herring 1999). Second, these heightened perceptions of risk cause investors to take actions to minimize their potential losses, that is, to withdraw capital from at-risk markets. In making these determinations investors consider a variety of factors including beliefs about political fundamentals and their influence on the likelihood of crises. However, in so doing, investors will privilege recent information, that is, information they have gleaned from the recent crisis. It is the revisions of these subjective assessments of political risk which, in turn, fuel investment decisions. These revisions are likely to be pronounced given the recency and salience of the crisis, both of which contribute to a propensity to privilege current information over prior beliefs in decision-making processes (Tversky & Kahnemen 1973).

Therefore, deviating from the current political economy literature, I contend that investors do not have fixed beliefs as to the efficacy of certain political environments. Rather, these dispositions are fluid and heavily conditional on information provided by the recent crisis. In effect, the crisis induces a discontinuity in their beliefs about the risk of particular political fundamentals or characteristics. That is, while investors may have some prior beliefs about over political fundamentals, the crisis itself disrupts these and causes investors to reach new subjective understandings as to the stability of the political climate. I argue that these updated beliefs, in turn, precipitate capital outflows and trigger self-fulfilling crises. Therefore, I anticipate that states with similar political fundamentals should be more likely to experience financial crises contagion. A proposition I test in the following sections.

4.2 RESEARCH DESIGN

To test the preceding proposition, I estimate a series of spatially and temporally lagged probit models of financial crises using maximum simulated-likelihood (MSL) by recursive-importance-sampling (RIS) as detailed in [chapter 2](#). Spatial estimation allows us to explicitly model and discriminate between the possible channels of contagion along which financial crises may be transmitted. In particular, MSL-by-RIS improves upon existing spatial work in the conditional crisis literature by accounting for the endogeneity in the relationship between the propensity for financial crises, that is, the risk of a crisis in country i is a function of the risk of a risk in country j and vice-versa.¹³ As noted in [chapter 2](#) and [Franzese *et al.* \(2014\)](#), failing to properly account for this simultaneity has been shown to produce biased estimates, and therefore our understanding of the contagion of financial crisis remains limited.

The characteristics of the model are given in [Equation 2.8](#) - [Equation 2.10](#), where y_{it} is the binary outcome, *Financial Crisis*, which takes the value of 1 for any country-year in which a banking crisis occurs are 0 otherwise, evaluated for all states between 1970 and 2007. The data comes from a new systematic database on financial crises from the IMF which identifies a banking crisis as having taken place when the financial sector experiences a large number of defaults, financial institutions face difficulties repaying contracts on time, non-performing loans increase sharply and all or most of the aggregate banking system capital is exhausted [Laeven & Valencia \(2012\)](#). The possible channels of contagion are included in the spatial filter $(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}$ as \mathbf{W} . Before detailing each of these channels, the general strategy is to first estimate a series of single- \mathbf{W} (e.g. one contagion channel) models, allowing me to better compare these to previous work (with the differences arising primarily from the proper estimation of the model). Then, to discriminate between these sources, I estimate a multi-parameteric model as discussed in [??](#). This entails a simple expansion of the spatiotemporal multiple to include several measures of spatial dependence:

¹³To my knowledge, only [Novo \(2003\)](#) has estimated a spatial lag model via MSL-by-RIS (or any analogous approach) on a single cross-sectional analysis of the ERM crisis.

$$y_{it}^* = (\mathbf{I} - \rho_1 \mathbf{W}_1 + \rho_2 \mathbf{W}_2 + \dots + \rho_R \mathbf{W}_R + \phi \mathbf{L})^{-1} + \mathbf{X}_{it} \boldsymbol{\beta} + \mathbf{u}, \quad (4.1)$$

with $\mathbf{u} = (\mathbf{I} - \rho_1 \mathbf{W}_1 + \rho_2 \mathbf{W}_2 + \dots + \rho_R \mathbf{W}_R + \phi \mathbf{L})^{-1} \boldsymbol{\epsilon}$

In order to test my argument on the importance of the similarity in political fundamentals for the contagion of crisis, I include a weights matrix of political proximity, $W_{PolProx}$. Specifically, ‘political proximity’ is measured as the inverted difference in the political constraints index [Henisz \(2000\)](#). Henisz’ measure is a spatial model (of a different sort) of political interaction – ranging from 0 to 1 – which identifies the number of veto players in a state and the distribution of their policy preferences. In effect, this measure captures the predicted difficulty of enacting policy reform, with each element of the weights matrix $W_{PolProx}$ representing the similarity of states in their ability to achieve such innovations. [MacIntyre \(2001\)](#) and others have argued that investors specifically look to this dimension when determining their response to a crisis, as such it is uniquely suited for my analysis.¹⁴ If my hypothesis is correct, then $W_{PolProx}$ should be positive and significant, suggesting the importance of political proximity in the spread of financial crises.

Prior work has included this measure, or its quadratic, as a right-hand side regressor in analyses of financial crises, so it is important to note what is distinct about my use of this measure. Specifically, the ‘political proximity’ spatial lag – and also the economic proximity spatial lag I discuss shortly – uses an associational or characteristic-based measure of proximity between a group of actors, traits that could (and have) been included separately as right-hand side regressors. In a model with a regressor and a spatial lag generate from the same underlying factor the parameters capture different aspects of the measure and, therefore, test fundamentally different theories. In particular, the right-hand side regressor tests whether there is a direct (non-spatial) effect of being that ‘type’ (e.g., whether states with fewer political constraints experience crises more frequently) on the outcome of interest, whereas the spatial lag tests whether there is a peer effect of being in that group (e.g., whether

¹⁴As an additional robustness check, I also estimate models in which political proximity is captured by the similarity of a states political institutions. This allows me to test the idea that investors have low information on political fundamentals and take cues from the most obvious signals (e.g., regime type). My results remain consistent to this alternative specification.

states with fewer political constraints experience crises concurrently).¹⁵ Spatial lags which use associational or characteristic-based measures of proximity between a group of actors offer us a means of capture indirect *and* potentially time-varying effects of group inclusion.

That is, even if we find no average direct effect of a group variable - via the coefficient on the right-hand side regressor - we would be incorrect in interpreting that directly as indicating that there is no effect of that group or type. Instead, it may suggest that the impact of being in a group of actors on the outcome is not constant across time (e.g., oscillating effects) and, therefore, has no significant average direct effect. However, with characteristic-based spatial lags, we are able to assess whether being part of a group itself causes actors to “move together” in that outcome or state across time.¹⁶ As such, they are uniquely able to test the theory I outline above over the evolving state of investor beliefs as to what constitutes politically risky/safe institutions. I make no claim about the likely direct effect of political institutions (captured by *PolConvV*), and defer to the previous literature on this point. My claim is only that states at a similar location on this measure (i.e., those with similar institutions) will have a higher probability of experience similar outcomes as a consequence of investor “lumping,” if correct this would be evidenced by a significant parameter estimate on $W_{PolProx}$.

In addition to my main measure, I also include three weights matrices to assess the common economic connections advanced in the literature. First, I include a matrix of geographic contiguity W_{Contig} which captures whether states share a common border, as indicated by the Correlate’s of War Direct Contiguity data (Stinnett *et al.* 2002). This both tests extant arguments on the importance of regional proximity in the spread of financial crises and proxies for a number of omitted spatially proximity features, ensuring the remaining spatial

¹⁵Perhaps an additional example will help to clarify this further. Say one was interested in how European Union (EU) membership influence growth. one could ask whether EU member states have higher growth rates than non-members, which could then be tested with a binary right-hand side regressor. Alternatively, one could ask whether the growth rates of the EU member states trend together, which could then be tested using a spatial lag with EU membership as the weights.

¹⁶There are, of course, other ways of capturing time-dependent effects of regressors such as regime-switching models. However, this strategy may be helpful in that we are not required to make strong claims about the periods in which we expect the regressor to have different impacts. Rather, we merely wish to hypothesize about about the co-movement of particular groups with respect to the outcome.

regressors are not erroneously capturing the effect of some omitted variable.¹⁷ As such, it is an extremely conservative estimation strategy, requiring the other spatial lags to have an effect which does not operate through simple geographic proximity. Second, to assess the impact of real spillovers, I include a matrix of trade flows W_{Trade} between all countries using the Correlates of War data on bilateral trade (Barbieri *et al.* 2009; 2012). Finally, I capture the importance of macroeconomic similarity by using the inverted difference in the level of development $W_{EconProx}$ with higher values representing ‘closer’ proximity.¹⁸

In addition to the possible channels of contagion, I include a number of additional country-level macroeconomic and political characteristics which are argued to increase the risk of financial crises (Gorton 1988, Kaminsky & Reinhart 2000). $GDPpc$ and $Population$ (both logged) are included to capture elements of the general socioeconomic climate.¹⁹ Additionally, government consumption as a share of GDP (Kg/GDP), investment as a share of GDP (Ki/GDP) and the ration of the money supply to international reserves ($M2/Res$) are included as measures of macroeconomic fundamentals.²⁰ Furthermore, to account for possible changes in the domestic economy which may trigger financial crises, I included the growth rate of GDP ($GDP\ Growth\ \%$), the inflation rate ($Inflation$), and the real interest rate ($Interest\ Rate$). Finally, to capture the potential direct importance of domestic political institutions, I included a measure of political constraints (as explained above).

¹⁷This was a flaw of prior analyses and is particular important given that many of the factors of interest (e.g., trade, political institutions, etc...) often cluster in geographic space.

¹⁸The level of development is seen by many as a key feature for predicting where financial crises and financial crises contagion will occur, as the similarity in market positions exposes them to more frequent common shocks (Bekaert *et al.* 2005, Lagoarde-Segot & Lucey 2006). I also estimate models using government consumption as a percentage of GDP and money supply over reserves and obtain similar results.

¹⁹Data on GDP, population, GDP growth, government consumption and investment come from the Penn World Tables 7.1 (Heston *et al.* 2012)

²⁰Data on money supply over reserves, the inflation rate, and the real interest rate all come from the IMF’s International Financial Statistics.

4.3 RESULTS

The results from the models of a single spatial lag indicate substantial support for the common channels of contagion (Table 4.1). Contiguity (Model 1), trade flows (Model 2, and economic proximity (Model 3) are all found to positively and significantly related to banking crises. That is, a banking crisis in one state significantly increases the risk of a banking crisis occurring in geographically or economically connected states. Furthermore, I find support for the role of common political institutions in spreading crises (Model 4). This indicates that a banking crisis increases the risk of additional banking crises in states with similar political fundamentals, as captured here by veto players.²¹ However, given that many of these factors are likely to be correlated (e.g., trade flows are greater between neighboring states), we need to estimate a more fully specified spatial model to discriminate between their individual effects.

As such, I estimate a multiparametric spatial lag model in which all of the spatial lags discussed previously are jointly estimated (Model 5). In these results, the impact of political proximity remains positive and significant even after accounting for other channels of spatial contagion. Moreover, the size of the dependence parameters ‘rho’ is quite high, suggesting the strength of the relationship and lending support for the role of an indirect political channel across which financial crises spread. We also observe that after controlling for the spatial-effect of political proximity there is no support for a direct (i.e., non-spatial) effect of political constraints on the realization of crises; the coefficient on *PolConV* is *insignificant*, or at least not one that is systematic over time.

Yet, the estimate of the spatial lag for political proximity suggests that these institutions *do* matter, simply not in the way that has been traditionally assumed. Instead, the evidence here suggests that states with similar political institutions are correlated in their

²¹In additional models (not reported), I also find support for this channel using less sophisticated measures of political similarity - reflecting the possibility of less-informed investors updating on basic cues - such as the Polity IV measure of democracy.

Table 4.1: Bank Crisis Contagion (Spatial Probit), 1970 – 2007

	Model 1	Model 2	Model 3	Model 4	Model 5
	Contig	Trade	Econ Prox	Pol Prox	Full Model
<i>GDPpc (logged)</i>	0.0377* (0.0196)	0.0405* (0.0197)	0.0385* (0.0199)	0.0521* (0.0247)	0.0484* (0.0246)
<i>Population (logged)</i>	-0.0026 (0.031)	-0.0045 (0.0314)	0.012 (0.0299)	0.0313 (0.0389)	0.0225 (0.0391)
<i>GDP Growth (%)</i>	- 4.1464*** (0.6337)	- 3.9939*** (0.6391)	- 4.1566*** (0.6328)	- 4.2302*** (0.8338)	- 4.1552*** (0.827)
<i>Kg/GDP</i>	0.0033 (0.0038)	0.0029 (0.0039)	0.003 (0.0038)	0.006 (0.0055)	0.006 (0.0055)
<i>Ki/GDP</i>	-0.003 (0.0031)	-0.002 (0.0031)	-0.0023 (0.0031)	-0.001 (0.004)	-0.0008 (0.004)
<i>M2/Reserves</i>	-0.0002 (0.0004)	-0.0002 (0.0003)	-0.0002 (0.0004)	-0.0002 (0.0005)	-0.0002 (0.0004)
<i>Inflation</i>	0.0000 (0.0001)	0.0000 (0.0001)	0.0000 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)
<i>Interest Rate</i>	0.0009 (0.0012)	0.0011 (0.0012)	0.0009 (0.0012)	0.0043* (0.0026)	0.0044* (0.0026)
<i>PolConV</i>	0.0074 (0.1157)	0.0519 (0.1157)	-0.0132 (0.1139)	0.0189 (0.1396)	0.03 (0.1396)
<i>LDV</i>	0.5170*** (0.0277)	0.5213*** (0.0258)	0.5216*** (0.0256)	0.5288*** (0.0324)	0.5236*** (0.0334)
<i>W_{Contig}</i>	0.0684*** (0.0258)	-	-	-	0.0588* (0.0372)
<i>W_{Trade}</i>	-	0.1703** (0.0567)	-	-	0.0665 (0.0755)
<i>W_{Econ Prox}</i>	-	-	0.1291** (0.054)	-	-0.0441 (0.0772)
<i>W_{Pol Prox}</i>	-	-	-	0.3569*** (0.0505)	0.2990*** (0.1064)
<i>Constant</i>	- 0.9609*** (0.3173)	- 0.8221** (0.3239)	- 0.9682** (0.3121)	- 1.0051** (0.4154)	- 0.9006* (0.4239)
<i>N (states)</i>	2459 (115)	2459 (115)	2459 (115)	2459 (115)	2459 (115)

SEs in parentheses. ***significant at 1%; **significant at 5%; *significant at 10%.

realization of financial crises, trending together within “types.”²² Thus, the evidence sug-

²²As with most of the work in empirical IR, more research is required to provide the micro-level support for my presumed casual mechanism. In future work I intend to estimate models using higher frequency data (e.g., portfolio flows) to more acutely test the behavior of investors in response to financial crises.

gests that no particular institutional design is unconditionally superior in the avoidance of crises. Rather, the impact of institutions appears to vary across, conditional upon whether crises have recently occurred in similar political environments. This possibility has not been previously explored in the literature, in part because it is a finding which is not possible to obtain from traditional non-spatial models.

Furthermore, the multiparametric model casts doubt on the role of direct economic interdependence in the spread of crises. Specifically, after accounting for other elements of spatial proximity, the effect of trade washes out and it is no longer found to be a significant channel for the contagion of crises. That is, the results suggest that the previous findings on the influence of trade on contagion may have been spurious. While potentially surprising for empirical researchers of contagion, this finding is actually more consistent with a number of case-study analyses which have frequently challenged the purported importance of the trade mechanism. Examining all crises since 1980, [Kaminsky *et al.* \(2003\)](#) identifies repeated episodes in which the impact of trade as a contagion mechanism was negligible. Additionally, [Athukorala & Warr \(2002\)](#) and [Dungey *et al.* \(2006\)](#) find little evidence of trade flows significantly contributing to the spread of the Asian financial crisis or the Mexican peso crisis. Furthermore, trade may have counter-veiling effects depending on the degree of concentration, with high levels making states more exposed to shocks, but export diversification making states better able to recover from them. A possibility which needs to be explored more deeply in future research.

To highlight the difference between the estimation strategy utilized here and prior work, Table 4.1) includes the results from commonly used alternatives (e.g., non-spatial probit and ‘naïve’ spatial probit).²³ The results evidence the importance of both: 1) including spatial lags and 2) properly accounting for their possible endogeneity. In particular, several of the country-level characteristics which are significant in the non-spatial probit model (Model

²³As a reminder, the ‘naïve’ estimator is one in which the spatial lages are simply included as exogenous right-hand side regressors and then estimated as in standard probit. In both the non-spatial and ‘naïve’ case temporal dependence is capture through the inclusion of the observed-lag of y (as opposed to latent- y^*), that is, a regime-switching model. I prefer this to other standard approaches – e.g., event-history based approaches such as counter and splines – because the regime switching model more directly captures the auto-dependence in y and therefore provides a more reasonable alternative to my preferred estimation strategy.

6) “lose” their significance when the spatial lags are included, as inflation and political constraints are no longer significant. More specifically, the effect of these regressors is now decomposed into an immediate non-spatial effect and the (long-run) spatial (and temporal) effect.²⁴ This casts further doubt on contemporary understandings of the role of political institutions in financial crises, and suggests that those theories advocating of their direct effect need to be reconsidered and evaluated using estimation strategies which account for spatial dependence.

While the ‘naïve’ spatial model offers some improvements over the non-spatial model in this respect, it is shown to drastically overestimate the impact of the spatial effects.²⁵ Significantly, it generates false positives for both trade and economic proximity, which find no support with the consistent estimator. Moreover, the estimated impact of contiguity and political proximity, while rightly signed and significant, are noticeably reduced when the appropriate estimator is employed. In all, the results suggest that previous research in economics which has treated the spatial lag as exogenous is likely to have overestimated the extent to which (some) ties matter.

As always, parameter estimates of non-linear models cannot be directly interpreted as substantive effects. As I noted in [chapter 2](#) and [chapter 3](#) these complications are even greater with the spatial models employed here given the interdependence in the outcomes. Therefore, using the approach identified in [chapter 2](#) I estimate conditional counter-factual effects via simulation. Using this strategy, I explore the extent to which the so-called “Tequilla crisis” in Mexico increased the risk of a banking crisis in neighboring Guatemala. In 1994, Mexico experienced a sudden devaluation of the peso, which triggered an economic crisis during which the stock market and banking system collapsed. These effects were far reaching, as additional crises soon followed in Argentina and Chile. This invites the question, what would the risk of crises in surrounding countries have been if not for the Mexican crisis? The answer to this question depends, in part, on the nature of the spatial relationship between Mexico and any other state. In [Figure 4.4](#), I depict the connections (the edges) between

²⁴This is analogous to the well-known impact of including a time-lagged dependent variable in one’s analysis

²⁵This confirms prior work which has found an upward bias in ‘naïve’ spatial probit models via monte carlo analysis ([Franzese *et al.* 2014](#)).

Table 4.2: Variation in Estimation Approaches

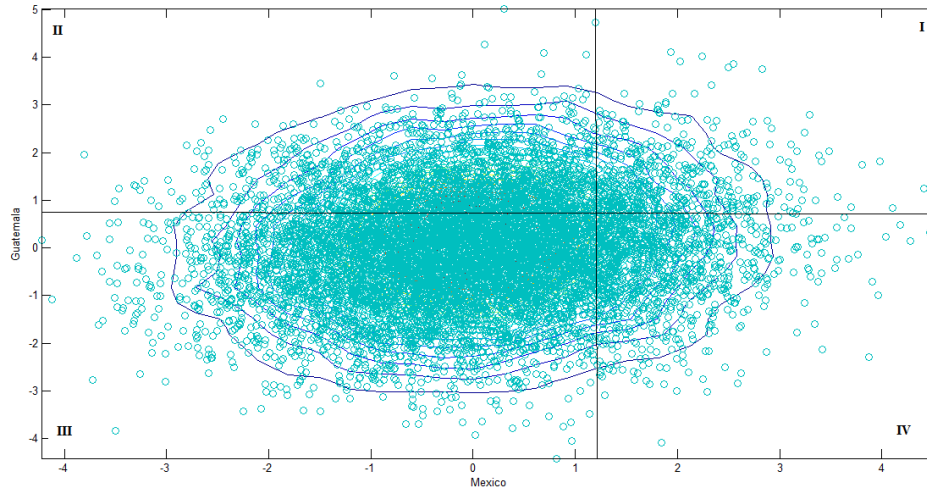
	Model 5	Model 6	Model 7
	MSL-by-RIS	Standard Probit	Naïve S-Probit
<i>GDPpc (logged)</i>	0.0484* (0.0246)	0.0520* (0.0350)	0.0685* (0.0361)
<i>Population (logged)</i>	0.0225 (0.0391)	-0.0634 (0.0570)	0.0440 (0.0592)
<i>GDP Growth (%)</i>	- 4.1552** (0.6337)	- 4.2706*** (0.6391)	- 3.725*** (0.6328)
<i>Kg/GDP</i>	0.0060 (0.0060)	0.0065 (0.0065)	0.0033 (0.0033)
<i>Ki/GDP</i>	-0.0008 (0.0040)	-0.0014 (0.0058)	-0.0008 (0.0061)
<i>M2/Reserves</i>	-0.0002 (0.0004)	0.0001 (0.0004)	0.0002 (0.0004)
<i>Inflation</i>	0.0001 (0.0001)	0.0002* (0.0001)	0.0002* (0.0001)
<i>Interest Rate</i>	0.0044* (0.0026)	0.0062* (0.0033)	0.0052* (0.0034)
<i>PolConV</i>	0.0300 (0.1396)	-0.3535* (0.1996)	-0.3475* (0.2060)
<i>LDV</i>	0.5236*** (0.0334)	-0.5960*** (0.0447)	-0.5563*** (0.0488)
<i>W_{Contig}</i>	0.0588* (0.0372)	-	0.7181*** (0.1995)
<i>W_{Trade}</i>	0.0665 (0.0755)	-	0.9035** (0.4262)
<i>W_{Econ Prox}</i>	-0.0441 (0.0772)	-	-1.0158** (0.4559)
<i>W_{Pol Prox}</i>	0.2990*** (0.1064)	-	1.8256* (0.9725)
<i>Constant</i>	- 0.9006*** (0.4239)	- 0.1033 (0.5828)	- 0.5246 (0.5090)
<i>N (states)</i>	2459 (115)	2459 (115)	2459 (115)

***significant at 1%; **significant at 5%; *significant at 10%.

Mexico in 1994 and all other countries in the sample (the nodes) for that year. Wider edges (e.g., connections) represent greater political similarity to Mexico, while ‘black’ colored edges indicate geographic contiguity.

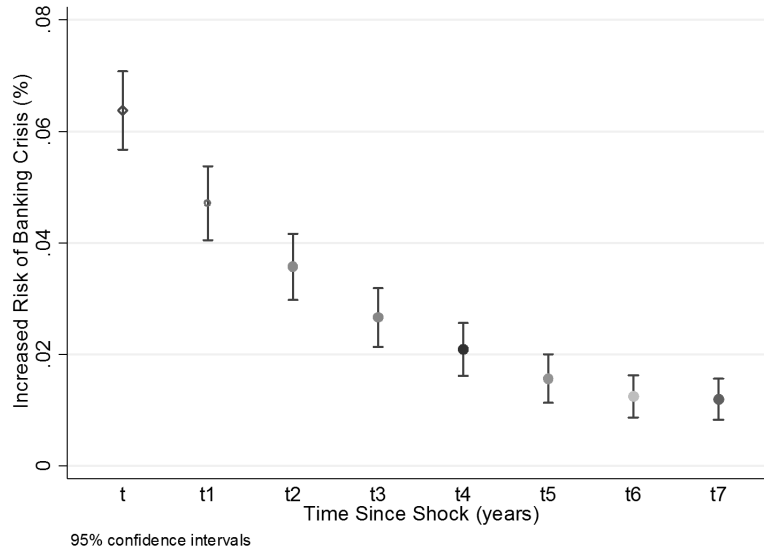
are -0.843 and -1.157 respectively. Therefore, a financial crisis occurs if the reduced-form disturbances are greater than 0.843 or 1.157. The distribution of the 10,000 disturbances in relation to the (adjusted) cutpoints is presented in Figure 4.5.

Figure 4.5: Scatter Plot of Reduced Form Disturbances (Guatemala & Mexico 1994))



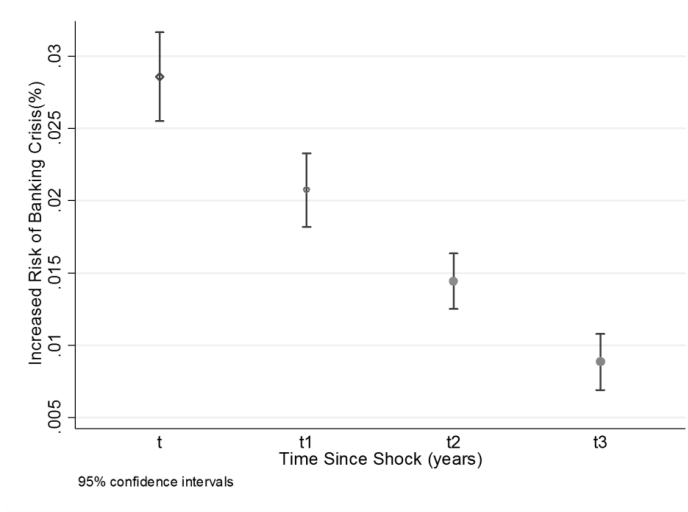
In these simulations Guatemala experiences a financial crisis 21.8% of the time when Mexico does not (1803/8268), yet 28.6% of the time when Mexico is undergoing a financial crisis (495/1732). Thus, the Mexican peso crisis increased the risk that the Guatemala would experience a similar crisis by 6.8%. To calculate our uncertainty about these estimates e.g., uncertainty over the estimates of the coefficient values we can resample the model parameters (β, ρ, ϕ) from a multivariate normal using the estimate means and variance-covariance from the preceding estimation. Doing so (100 times) indicates that with 95% confidence the immediate effect estimate lies between 5.67% and 7.07% (with a mean of 6.37%).

Figure 4.6: Short and Long Run Effect of Shock' (Guatemala and Mexico, 1994–2000)



However, given the time-series nature of the data, we are also interested in calculating the long-run counter-factual effects. That is, what effect did the 1994 Mexican peso crisis have on the risk of a financial crisis in the Guatemala in the years which followed? To calculate this we draw 10,000 i.i.d standard normal disturbances, as before, and compare them to their reduced-form cutpoints. After calculating the instantaneous effect, we group those 10,000 realizations of the data into two groups, one in which Mexico experienced a financial crisis (state of the world A) and the other in which they did not (state of the world B). Next, we calculate whether the Guatemala was above or below its cutpoint in subsequent years, that is, whether the Guatemala experienced a financial crisis in $t + 1, t + 2, \dots, t + n$. Finally, we calculate the difference in the estimated frequencies across groups A & B. This difference represents the long-run effect of that initial shock (e.g., the Mexican peso crisis).

Figure 4.7: Short and Long Run Effect of ‘Shock’ (US and Mexico, 1994–1997)



The distribution for the estimated effects of all 100 trials is given in Figure 4.6 calculated up to $t + 3$ with 95% confidence intervals. As expected the immediate effect is strongest, 6.37%, with diminishing, but still significant, effects in subsequent years: 4.71%, 3.57%, 2.66%, 2.09%, 1.57%, 1.25%, and 1.19%. Thus, the Mexican peso crisis significantly increased the risk of a crisis in the Guatemala not simply in 1994, but also in several years to follow. Moreover, these effects appear highly persistent as we fail to return to a steady-state (no difference) by 2001.

This contagious impact is not exclusive to small markets, but afflicts large markets as well. For example, the risk of the United States experiencing a financial crisis was 2.85% [95% between 2.54 % and 3.61%] greater when Mexico underwent a financial crisis in these simulations.²⁶ Furthermore, the effects persisted for several years following the initial crisis. Though the immediate effect is the strongest (likely via contiguity which proxies for a number of different kinds of connections), significant effects remain in subsequent years:

²⁶Relative to Guatemala, the risk of the US undergoing a banking crisis at all is significantly lower – as one might expect – occurring on average in around 5% of the simulations. Thus, the percent increased risk is quite a substantial effect (more substantial than the Guatemalan case).

2.07%, 1.44%, and 0.88% (see Figure 4.7). Ultimately, the difference between the two groups converges to zero as their respective likelihoods of experiencing a crisis return to a common steady-state probability. These effects may explain President Clinton's rationale behind extending Mexico a highly unpopular \$20 billion assistance package to Mexico following the initial crisis. The rationale for and impact of such interventions is a topic I intend to explore in future work.

4.4 DISCUSSION

Increased global integration – political and economic – offers considerable benefits, but also carries substantial risk. In so many words, this sentiment reflects the two stylized facts which underpin much of the contemporary thoughts on the consequences of interdependence. One of the major downside risk of such integration is the possible contagion of financial instability between states, as evidenced most dramatically by the recent ‘Great Recession.’ Such events highlight and need to improve our understanding of the process by which crises spread internationally and what, if anything, states can do to minimize these risks?

In this chapter, I focus on the first half of this question, examining the factors which increase the risk of financial crisis contagion. My analysis suggests that the presumed importance of direct spillovers (such as trade) on contagion may be overstated, as it lacks support when other spatial factors are accounted for. However, these results also suggests that political factors may play a more important, and different, role than previously thought. In particular, I find that states with similar domestic political institutions are likely to correlate highly in their realization of crises, that is, countries with ‘closer’ institutions have a greater risk of financial contagion. This departs from traditional political economy theory which has exclusively focused on the direct role of political institutions in the onset of crises. Instead, I argue for an additional indirect role of political institutions wherein countries with similar political fundamentals are more likely to experience crises concurrently. That is, we observe contagion amongst similar types of political regimes.

I argue that this spread occurs as investors update their beliefs on the risk posed by particular political environments following a crisis. From these updated beliefs, they withdraw capital from risky environments, therein generating self-fulfilling crises. This highlights the importance in understanding now just how states are connected explicitly (or directly), but how they are *perceived* to be connected which constructs a tie between states. As the old saying goes, perception becomes reality. Though in this version of the project I do not test the actions of investors directly, such analysis is planned in the future to provide additional support for the mechanism suggested here. Specifically, by using higher frequency financial data – such as portfolio investments – we can model the short-term behavior of investors in response to a crisis directly. If in this alternative analysis we also observe the effect of political proximity, it would provide strong evidence for the mechanism advanced here.

As alluded to previously, an additional avenue for future research concerns the policy prescriptions available to states which may minimize or mitigate the risk of contagion. In particular, whether financial interventions – such as the example of the United States in Mexico – help to stymie the spread of economic instability. Current research offers tentative support for this effect, yet this work has consistently underestimated the role of contagion in spurring such support. That is, the relationship between financial contagion and intervention runs in two direction, with two related questions: does the anticipation of contagion increase the likelihood of intervention *and* does intervention reduce the risk of contagion? The existing scholarship on these issues does not consider them jointly and, as such, risk a biased understanding of these effects. Therefore, in planned work, I attempt to isolate the individual effect of intervention to improve our understanding of whether states can act to reduce the downside risk of interdependence.²⁷

²⁷Data constraints currently prevent a full analysis of this type, though new data on bilateral financial bailouts in [Schneider \(2013\)](#) should facilitate analysis of this type when publicly available.

5.0 UNOBSERVED HETEROGENEITY AND RARE EVENTS

As has been noted elsewhere, many of the most important phenomena in international relations are both discrete and rare events (King 2001, King & Zeng 2001b). That is, binary outcomes where the number of historical occurrences (i.e., ones) of the event in question is extremely low, both in absolute terms and relative to the number of historical non-occurrences.¹ Much of the work in peace studies focuses on events that are rare, including the initiation of and entry into international and civil wars, attempts to overthrow the government via rebellion and/or military coups, and the use of genocide by state actors to suppress such efforts. However, work in IPE and IOs also regularly deals with rare events: the formation of preferential trading agreements and bilateral investment treaties, the decision to employ economic sanctions, the initiation of processes of adjudication within international dispute bodies, the onset of banking and currency crises. These are some of the most significant phenomena within IR scholarship.² Yet, researchers in these subject areas continue to neglect the way in which the rarity of these events may be biasing their analyses.

Therefore, in this chapter, I review some of the more significant problems which arise in estimating models of rare-event data. In particular, two issues have been raised for models with sparse data. First, with low event totals maximum likelihood estimates are known to be biased, as the conditional density of the regressors about the less frequent outcome

¹Prior work has defined rarity exclusively as the proportion of events to non-events (King & Zeng 2001b). However, the inferential problems presented from rare events is a consequence of a small sample of occurrences - in effect, a small sample bias - and not as directly a consequence of the percentage of the occurrences within the sample.

²Of course, such events are not unique to IR and examples abound in American and Comparative Politics as well.

is poorly defined. Second, with sparse data complete or quasi-separation is more likely to arise, producing infinite-valued estimates for these parameters. It is this second concern which, in part, has led some researchers to caution against the use of fixed effects with rare event data, as they states which never experience the outcome are dropped from the analysis [Beck & Katz \(2001\)](#). While both problems are familiar to political science, they have been previously been cast as separate concerns. In the next section, I indicate how they are both the result of the same underlying issue: small sample bias.

After recognizing and clarifying this, I propose a simple and general solution, Penalized Maximum Likelihood (PML), which addresses both of these issues. Building on the initial formulation given by [Firth \(1993\)](#), penalized likelihood has become an increasingly common strategy for removing the first-order bias present in maximum likelihood. Departing from previous bias-reduction strategies, [Firth \(1993\)](#) proposed introducing a slight penalty to the score function – equivalent to Jeffreys invariant prior in the case of logistic models – to reduce the small sample bias. Importantly, this approach does *not* rely on first obtaining estimates of the parameters, as other correction-based approaches do. As a result, PML is able to provide unbiased estimates for parameters even in instances where ML is unable to produce stable estimates, such as in cases of quasi- or complete separation. Therefore, PML provides clear advantages over both conventional rare-events pooled or standard panel estimators. Specifically, it allows researchers to estimate models which both: *i*) correct for the small sample bias induced from rare events and *ii*) include fixed effects with none of the attendant sample censoring which results from estimating conventional fixed effects models of rare events. [Cook et al. \(2014\)](#) have argued that this second benefit, estimates for all the unit effects, is important in that it allows researchers to estimate substantive quantities of interest such as marginal effects which are not possible (or badly biased) with conventional fixed effects approaches.

While it is shown that penalized maximum likelihood fixed effects should often be preferred over available alternatives, there are two limitations to this approach. First, though the penalization strategy does reduce the incidental parameter bias, it may still be significant in samples with very small- t . Second, available programs for estimating these models are

ill-suited for sampling dimensions common to International Relations, which can include tens of thousands unit parameters (e.g., dyads). As such, I suggest an additional complementary group-fixed effects strategy. In short, while assuming a common propensity (i.e., pooled) is rarely supported, assuming complete heterogeneity (i.e., unit-fixed effects) is often equally unrealistic (and/or unnecessary). If sub-groups of the sample are sufficiently heterogeneous we can achieve the benefits of unit-fixed effects, while reducing the incidental parameter bias. Therefore, I conclude the chapter with a discussion of how we might identify these groups to allow for such estimation.

5.1 PENALIZED MAXIMUM LIKELIHOOD

While maximum likelihood estimation is typically motivated on the basis of its asymptotic fitness, it is well known to produce bias in finite samples. As I discuss below, because this bias is order $O(n^{-1})$ it does not usually constitute a problem for large samples, however in moderate to small samples, or those with low total Fisher information these biases can be potentially severe (Cordeiro & McCullagh 1991). In addition to biased estimates of the parameters, the asymptotic confidence intervals also often perform poorly (Heinze 2006). King & Zeng (2001a;b) introduced several of these concerns to political science, noting and clarifying that the bias of MLE estimates was not simply a function of its absolute size of the sample, but dependent on the number instances of the event in the data.³ It is for this reason that Heinze (2006) notes that studies using ML with extremely sparse data are not to be trusted.

The nature of this bias for the logistic model has received particular attention (Cordeiro & McCullagh 1991, Firth 1993, Schaefer 1983).⁴ Recall the now familiar latent variable representation given in Equation 1.1 and Equation 1.2, the probability that $y_{it} = 1$ is given by:

³Or, more generally, the number of instances of the less frequent outcome.

⁴See Schaefer (1983) for an analytic expression of the bias

$$Pr(y_{it} = 1|x_{it}) = \pi_{it} = (1 + \exp\{-\mathbf{x}_{it}\boldsymbol{\beta}\})^{-1} \quad (5.1)$$

The parameters are then estimated via by maximizing the log-likelihood:

$$\begin{aligned} \ln L(\boldsymbol{\beta}|y) &= \sum_{\{Y_{it}=1\}} \ln(\pi_{it}) + \sum_{\{Y_{it}=0\}} \ln(1 - \pi_{it}) \\ &= \sum_{i=1}^N \ln(1 + \exp\{(1 - 2Y_{it})\mathbf{x}_{it}\boldsymbol{\beta}\}) \end{aligned} \quad (5.2)$$

Which produces $\hat{\boldsymbol{\beta}}$, an asymptotically unbiased estimator of $\boldsymbol{\beta}$. However, in finite samples it is biased. The asymptotic value of the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ is given by the following expansion:

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + \frac{b_1(\boldsymbol{\beta})}{n} + \frac{b_2(\boldsymbol{\beta})}{n^2} + \frac{b_3(\boldsymbol{\beta})}{n^3} + \dots$$

Therefore, as n increases, or as more information is added, the accuracy of our estimate of $\boldsymbol{\beta}$ grows. A variety of strategies have been proposed to eliminate the first-term of this expansion, $\frac{b_1(\boldsymbol{\beta})}{n}$, thereby producing first-order $O(n^{-1})$ unbiased estimates. [King & Zeng \(2001b\)](#)'s offer one such strategy for estimating this $\frac{b_1(\boldsymbol{\beta})}{n}$, which derives from [Cordeiro & McCullagh \(1991\)](#)'s weighted least-squares expression:

$$\frac{b_1(\boldsymbol{\beta})}{n} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\boldsymbol{\xi}$$

Where $\boldsymbol{\xi}_i = 0.5Q_{ii}[(1 + w_1)\hat{\pi}_i - w_1]$, Q_{ii} are the diagonal elements of $\mathbf{Q} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{W} = \text{diag}\{\hat{\pi}_i(1 - \hat{\pi}_i)w_i\}$. Estimating and subtracting the bias from $\hat{\boldsymbol{\beta}}$ yields $\tilde{\boldsymbol{\beta}}$, a nearly unbiased estimate of $\boldsymbol{\beta}$. This approach is similar in spirit to a range of similar bias-corrective strategies presented in the broader literature on ML-estimation. [King & Zeng \(2001a;b\)](#) show the benefits from bias-correction over conventional 'pooled' logit, particularly for small samples and rare events.

However, additional problems arise from sparse data – e.g., rare events – when we shift focus to panel estimation. As discussed in [chapter 1](#), the most reasonable assumption for re-BTSCS International Relations data is often that there are unobserved unit effects which are correlated with the included regressors (e.g., α_i and \mathbf{x}_{it} are correlated), implying the need for fixed effects. One strategy for estimating fixed effects models, ‘unconditional logit,’ is to include additional dummy variables for each unit (but 1) into our model and maximize the standard logit log-likelihood. However, this suffers from the well-known incidental parameters bias and is inconsistent in N when T is fixed.⁵ In short, for any panel with fixed- T estimators including individual fixed effects (i.e., nuisance parameters) provide inconsistent estimates of the structural parameters of interest (Neyman and Scott 1948).⁶ As such, Chamberlain (1980) proposes an alternative estimator, ‘conditional logit,’ using a conditional likelihood function, conditioning on a set of sufficient statistics ($\Sigma_t y_{it}$) for the incidental parameters (α_i).⁷ That is, concentrating out the fixed effects. As opposed to unconditional logit, conditional logit does produce estimates which are consistent in N . However, either strategy – conditional or unconditional – is problematic in the presence of rare event data as those units which do not change state – e.g. $\Sigma_t y_{it} = 0$ or T – contribute nothing to the likelihood (returning estimates of $\pm\infty$) and are consequently dropped from the analysis.⁸ This issue has been widely noted and debated within political science, with researchers arguing that it induces a form of sample selection bias and therefore cautioning against the use of fixed effects [Beck & Katz \(2001\)](#).

Therefore, two issues have been raised in political science as concerns when estimating binary and rare event data. First, with increasingly rare events maximum likelihood estimates will likely be biased and asymptotic standard errors problematic ([King & Zeng 2001a;b](#)). Second, with rare events it becomes increasingly likely that some units will fail to

⁵Lancaster (2000) provides a useful summary of (and the history behind) the incidental parameters problem.

⁶More specifically, the estimates of the incidental parameters themselves are inconsistent because we have only a limited number of observations from which to estimate α_i (i.e., T). In turn, this inconsistency transmitted to the estimates of the structural parameters if we can not derive estimators which do not depend on the incidental parameters (Heckman 1981).

⁷Chamberlain (1980) builds on the work of Anderson (1973) to allow for multiple regressors.

⁸Whether they are literally ‘dropped’ from the estimation varies based on the statistical package one uses. Yet, in all the point remains that they are contributing nothing to the likelihood and, therefore, coefficient estimates of the regressors of interest.

experience the outcome within the sample period. Meaning researchers are forced to either improperly handle unit heterogeneity (e.g., estimate a 'pooled' or random effects model) or face sizable sample losses from estimating fixed effects models (Beck & Katz 2001). While these have traditionally been cast as separate concerns, the problems actually arise from the same fundamental issue: small samples. In the former there are too few events in the total sample, while in the latter there are too few events in the particular unit-sample (up to, and including, none). To be precise, the fundamental problem in each case is the lack of a sufficient balance of the two outcomes (0 and 1) over the distribution of some X . Consider, King & Zeng (2001b)'s intuitive illustration of the 'rare event' bias using a simple two variable model, showing that with too few realizations of $Y = 1$ the density for $Pr(X|Y = 1)$ will be calculated poorly and, in turn, the boundary condition incorrectly located. It is easy to see separation as a special case of this problem, where the number of observations of $Y = 1$ is zero and, as such, X – in this case the unit dummy – perfectly predicts the outcome.

While separation is rightly considered an extreme version of the small sample problem, we cannot use the strategy proposed by King & Zeng (2001b) in such situations. While King & Zeng (2001b)'s strategy *could* (possibly) suffice to remove the incidental parameter bias in unconditional fixed effects models for those units which experience at least one failure, it offers no assistance in obtaining estimates for those units which never experience the outcome. King & Zeng (2001b)'s approach, and other bias corrective strategies, relies on first obtaining finite estimates of the parameters in order to subsequently 'correct' the bias. However, as noted by Albert & Anderson (1984) and Santner & Duffy (1986) there are conditions under which such estimates will not exist, namely, quasi and complete separation. Therefore, bias-correction approaches offer us little recourse when facing such data. *However*, another bias reduction strategy has been proposed by Firth (1993) which instead modifies the score function to remove the bias during the maximization process itself. As such, it does not rely on the maximum likelihood estimates, making it instead what Firth (1993) calls a

bias ‘preventive’ approach.⁹ Accordingly, it is able to produce finite parameter estimates of parameters even in the presence of separation (Heinze & Schemper 2002, Zorn 2005).¹⁰

As noted above, this is achieved by a modification to the score function (which, as we will see, is equivalent to penalizing the likelihood by Jeffreys invariant prior):

$$U_r^*(\theta) = U_r(\theta) + A_r(\theta) \quad (5.3)$$

Where $U_r(\theta)$ is the ordinary score and $A_r(\theta)$ is modification to the score derived from the data. For instance, in the exponential family of models this adjustment is given by:

$$a_r = \frac{1}{2} \text{tr} \left\{ i^{-1} \left(\frac{\partial i}{\partial \theta_r} \right) \right\} = \frac{\partial}{\partial \theta_r} \left\{ \frac{1}{2} \log |i(\theta)| \right\} \quad (5.4)$$

With the solution of $U_r^*(\theta) = 0$ locating a stationary point of:

$$l^*(\theta) = l(\theta) + \frac{1}{2} |I(\theta)| \quad (5.5)$$

or...

$$L(\theta) = L(\theta) |I(\theta)|^{\frac{1}{2}} \quad (5.6)$$

That is, the ordinary likelihood L penalized by the square root of the determinant of the information matrix $|I(\theta)|^{\frac{1}{2}}$, which is equivalent to Jeffreys prior. For the logit likelihood:

$$L^*(\theta) = \prod_{i=1}^n \left(\frac{1}{1 + \exp(-x_i \beta)} \right)^{y_i} \left(1 - \frac{1}{1 + \exp(-x_i \beta)} \right)^{1-y_i} \quad (5.7)$$

⁹Other bias-preventive approaches have been suggested (see Kosmidis (2007) for a review) which make similar adjustments. However, these are (i) non-iterative, (ii) shrink toward the mean (rather than zero), and (iii) do not fully remove the first-order bias. Moreover, simulation studies provide evidence for the superiority of the modified score function approach over these possible alternatives (Heinze & Schemper (2002)). As such, I do not discuss these further.

¹⁰A point also made by King & Zeng (2001b) in their footnote discussing Firth (1993).

the determinant of the information matrix is maximized when $\beta = 0$, so the penalty function shrinks the estimates – including unit effect estimates that would otherwise be infinite – toward zero. Furthermore, penalized maximum likelihood should help ameliorate the small sample bias given that the attendant bias from the incidental parameters is (always) one of overestimation, the shrinkage imposed by penalization should provide superior estimates of the incidental and, in turn, structural parameters even in small samples.¹¹ In more general terms, it reduces the bias inherent to small samples in ML-estimation and, additionally, it reduces the incidental parameter bias in conditional fixed effects models, up to and including those units which never experience the outcome. As such, penalized maximum likelihood directly addresses *both* of the primary concerns voiced in International Relations regarding rare events estimation.

An additional benefit of the penalized maximum likelihood strategy is the flexibility of the estimator itself. Kosmidis & Firth (2009) develop or discuss bias-reduction adjustments for a broad class of generalized linear and nonlinear models (both with and without known dispersion parameters). Including a strategy for penalized maximum likelihood probit – using modified iterative reweighted least squares with data adjustments through pseudo-responses¹² – which presents an interesting possibility for panel binary-outcome data. Previously, researchers with panel binary-outcome data have been presented with two estimators: random effects and conditional fixed effects logit (Maddala 1987). Fixed effects probit (with dummies) is inconsistent in small samples and therefore cautioned against, and there is no sufficient statistic to condition on – as in the logit case – and therefore no conditional fixed effects probit estimator is available. Penalized maximum likelihood fixed effects presents an alternative which produces more accurate estimates using either link function, and therefore expands the set of options available to researchers with panel binary-outcome data. Given

¹¹Additionally, it will, with certainty, remove the first-order bias inherent to maximum likelihood estimation (especially salient for small samples), which is the chief utility of penalized maximum likelihood. Given this benefit, we would argue that penalized maximum likelihood (with or without fixed effects) should often be preferred to standard ML estimation.

¹²In the probit case the pseudo-responses are:

$$y^* = y - \frac{h\pi(1-\pi)\eta}{\{2\phi(\eta)\}}$$

that ultimately we want a model of dynamics *and* unit heterogeneity this is important as many of the developments for dependent binary-outcome models have been made in probit (see [chapter 2](#)). I discuss possible strategies for integrating these two approaches in [chapter 7](#).

5.2 RESULTS

Therefore, I explore the small-sample properties of the three non-panel estimators (i.e., pooled, rare-events logit, and penalized maximum likelihood) and four panel estimators (i.e., random effects, unconditional fixed effects, conditional fixed effects, and penalized maximum likelihood fixed effects) via Monte Carlo simulation.¹³ These experiments draw from and build upon the work in [Cook *et al.* \(2014\)](#), with many of the findings appearing in that text as well. The specification of the data-generation process is similar to that found in [Greene \(2004\)](#), with the basic framework for binary-TSCS data given in [chapter 1](#).¹⁴ In addition to the unit effect (α_i), the DGP includes a time-varying exogenous regressor (X_{it}^1), a time-varying endogenous regressor (X_{it}^2), a time-invariant exogenous regressor (X_i^3), and a time-varying endogenous dummy (d_{it}). More specifically, the DGP is:

$$y_{it} = \mathbf{1}[\beta_1 X_{it}^1 + \beta_2 X_{it}^2 + \beta_3 X_i^3 + \beta_4 d_{it} + \alpha_i + \epsilon_{it} > 0] \quad (5.8)$$

Where $i = 1, \dots, N$, $t = 1, \dots, T$ and...

$$X_{it}^1 \sim N(0, 1)$$

$$X_{it}^2 = \tilde{X}_{it}^2 + \phi\alpha_i, \text{ where } \tilde{X}_{it}^2 \sim N(0, 1)$$

$$X_i^3 \sim N(0, 1)$$

$$d_{it} = \mathbf{1}[\tilde{X}_{it}^4 + \phi\alpha_i > 0], \text{ where } \tilde{X}_{it}^4 \sim N(0, 1)$$

¹³I should note that attempts were also made to estimate rare-events logit with fixed effects, but the parameter estimates which results were badly biased (much worse than even unconditional fixed effects). As I currently have little intuition for their disproportionately poor performance, I choose to not include them here.

¹⁴I have also ran a series of experiments replicating the DGP utilized by [King & Zeng \(2001b\)](#) in their analysis of rare events. The results show, as expected, that the PML performs almost identically to re-logit under a variety conditions (when separation is not a concern). In the next iteration I plan to begin the results section with this discussion/results but simply did not have time at present.

while the fixed-effect (α_i) is varied across experiments, drawn from from one of three following distributions:

$$\alpha_i \sim N(-3, 1)$$

or...

$$\alpha_i \sim -\chi^2(3)$$

or...

$$f(\alpha_i) = p(g_1(\alpha)) + (1 - p)(g_2(\alpha))$$

where $g_1(\alpha_i) \sim N(0, 1)$, $g_2(\alpha_i) \sim N(-6, 1)$, and $p = 0.5$

That is, the performance of estimators are evaluated when the unit effects are distributed normally, χ^2 , and bimodal (with each peak being normally distributed).¹⁵ The random effects model assumes a parametric distribution for the unit effects. If this distributional assumption is false, it should negatively affect the estimators performance and non-parametric estimators (e.g., fixed effects) should perform better. To simulate rare events, the mean of the unit effects (for each distribution) is approximately -3, which sufficed to ensure that some censoring occurs in each of the experiments. The covariance parameter ϕ determines the degree to which the variables X_{it}^2 and d_{it} are endogenous – the extent to which they covary with the unobserved unit effects – which is also varied across experiments $\{0, 0.25, 0.5\}$.¹⁶ Throughout both N and T are also varied, resulting in a large number of experimental conditions, though for the sake of concision I primarily confine attention here to the $N = 50$, and $T = 20$ case.

In the first set of experiments, α_i is drawn from a normal distribution, with all $\beta = 1$, and varying ϕ between 0, 0.25, and 0.5. [Table 5.1](#) presents the results from the estimators which assume a common intercept: logit, rare-events logit, and penalized-logit. In experiment #1 – the top row set – ϕ is set to 0 and we see that, as expected, the results from each of the estimators is biased. This is consistent with [Wooldridge \(2010\)](#)'s finding that for probit models the omission of exogenous variables results in attenuation in the estimates of

¹⁵Figures of the representative probability density of each distribution are provided in the Appendix I.

¹⁶This also, of course, determines the degree of censoring in the data, with higher levels of endogeneity resulting in greater sample losses.

the included regressors.¹⁷ Cramer (2005) has argued that a similar bias exists for logit, and these results seem to offer further support for such claims. On average, we observe that the pooled estimators underestimate by about 10-12% of their true values. More surprisingly, among the set of pooled estimators, conventional logit appears to do relatively well, consistently outperforming the two bias-reductive approaches. This relative performance generally holds even as we add endogenous regressors in experiments #2 and #3, yet the nature of the bias switches as all three estimators now produce inflated estimates of the endogenous parameters degrades rapidly. In sum, these findings suggest that neglecting unit heterogeneity always biases the parameter estimates in logit models (with the direction of the bias dependent upon the covariance between the included regressors, the excluded regressors, and the outcome).

While one would not expect rare-events logit or penalized maximum likelihood to offer much improvement over conventional logit with respect to this model misspecification – as they too offer no accounting of the unit heterogeneity – it seems odd that they offer no improvement to the small sample bias which should also be present in the data. Believing that this was likely a consequence of the relatively moderate N -to- T ratio in the sampling dimensions, I reran experiment #3 with an expansion to N (to 100) and a reduction in T (to 2), to see whether the bias-reductive approaches would offer greater improvements under these conditions. While they do offer some benefits over the conventional logit estimator – most dramatically in the estimation of time-varying endogenous regressors – these gains are not on the order that we would expect to observe given the short T specification. As such, further work is necessary to determine the conditions under which bias-corrective approaches should be preferred. For now, I confine attention to standard-ML logit in any of the remaining discussion on pooled estimators.¹⁸

¹⁷However, Wooldridge (2010) shows that this does *not* affect the marginal effect estimates

¹⁸These are an admittedly constrained set of experiments to draw any definitive conclusions from. However, given the strength of the claims made by advocates of penalized maximum likelihood – e.g., Paul Allison has stated that “a case could be made for *always* using penalized likelihood” – it is important to understand the conditions under which this strategy may not result in improvements over traditional ML.

Table 5.1: Coef. Est. Pooled Models ($N=50, T=20, \alpha_i \sim N$, 1000 trials)

		ML-Logit	RE-Logit	PML-Logit
#1: $\phi = 0$	<i>TV_EXO</i> (β_1)	0.889 (0.114)	0.880 (0.113)	0.880 (0.113)
	<i>TV_END</i> (β_2)	0.881 (0.113)	0.872 (0.111)	0.872 (0.111)
	<i>TIV_EXO</i> (β_3)	0.881 (0.187)	0.872 (0.185)	0.872 (0.185)
	<i>END_DUM</i> (β_4)	0.880 (0.205)	0.882 (0.203)	0.882 (0.203)
#2: $\phi = 0.25$	<i>TV_EXO</i> (β_1)	0.908 (0.139)	0.900 (0.137)	0.900 (0.137)
	<i>TV_END</i> (β_2)	1.111 (0.135)	1.095 (0.132)	1.095 (0.132)
	<i>TIV_EXO</i> (β_3)	0.906 (0.208)	0.893 (0.205)	0.893 (0.205))
	<i>END_DUM</i> (β_4)	1.237 (0.268)	1.226 (0.263)	1.226 (0.263)
#3: $\phi = 0.5$	<i>TV_EXO</i> (β_1)	0.945 (0.171)	0.926 (0.166)	0.926 (0.167)
	<i>TV_END</i> (β_2)	1.267 (0.173)	1.241 (0.168)	1.241 (0.168)
	<i>TIV_EXO</i> (β_3)	0.935 (0.227)	0.915 (0.221)	0.915 (0.222)
	<i>END_DUM</i> (β_4)	1.505 (0.401)	1.496 (0.387)	1.496 (0.388)
N=100; T=2; $\phi = 0.5$	<i>TV_EXO</i> (β_1)	1.053 (0.469)	0.915 (0.391)	0.928 (0.398)
	<i>TV_END</i> (β_2)	1.415 (0.461)	1.236 (0.368)	1.252 (0.381)
	<i>TIV_EXO</i> (β_3)	1.018 (0.480)	0.888 (0.409)	0.898 (0.414)
	<i>END_DUM</i> (β_4)	1.569 (1.007)	1.508 (0.827)	1.518 (0.851)

Note: Standard deviations across trials given in parentheses. True β 's all equal to 1. X_{it}^2 is included in both the pooled and random effects models, but β_2 is not reported here.

As we have seen, when there is any unit heterogeneity in the data pooled estimators will be biased and inconsistent. How do the explicitly panel based estimators fare under these same conditions? The performance of these is given in [Table 5.2](#). As before, in experiment #1 ϕ is set to 0 and on average, across the simulations, observe eight units with no variation in the dependent variable (meaning several of the estimators will be based on different sample

sizes). Under these conditions the estimators perform roughly as expected. The random effects model reflects the true DGP – there is no correlation between ϕ and the regressors and distributional assumption is correct– and as such it performs quite well, producing the efficiency gains one would expect.

Table 5.2: Coef. Est Panel Models ($N=50, T=20, \alpha_i \sim N, 1000$ trials)

		(Random)	(Unc-FE)	(Con-FE)	(PML-FE)
#1: $\phi = 0$	<i>TV_EXO</i> (β_1)	1.011	1.095	1.012	0.997
		(0.128)	(0.145)	(0.130)	(0.128)
	<i>TV_END</i> (β_2)	1.005	1.089	1.006	0.991
		(0.122)	(0.139)	(0.125)	(0.123)
	<i>END_DUM</i> (β_4)	1.011	1.09	1.009	0.994
		(0.227)	(0.252)	(0.232)	(0.228)
#2: $\phi = 0.25$	<i>TV_EXO</i> (β_1)	1.011	1.106	1.013	0.992
		(0.156)	(0.182)	(0.162)	(0.141)
	<i>TV_END</i> (β_2)	1.116	1.109	1.016	0.970
		(0.148)	(0.171)	(0.152)	(0.140)
	<i>END_DUM</i> (β_4)	1.179	1.109	1.010	0.911
		(0.291)	(0.171)	(0.298)	(0.266)
#3: $\phi = 0.5$	<i>TV_EXO</i> (β_1)	1.125	1.014	1.024	0.964
		(0.223)	(0.185)	(0.195)	(0.179)
	<i>TV_END</i> (β_2)	1.120	1.232	1.019	0.959
		(0.209)	(0.179)	(0.183)	(0.168)
	<i>END_DUM</i> (β_4)	1.094	1.371	0.993	0.966
		(0.482)	(0.424)	(0.433)	(0.403)

Note: Standard deviations across trials given in parentheses. True β 's all equal to 1. X_{it}^2 is included in both the pooled and random effects models, but β_2 is not reported here.

Unconditional fixed effects is biased as well – suffering from the well known incidental parameter bias in samples of this size – producing estimates around 8-9% greater than the true values, which is roughly consistent with the results reported by Katz (2001) and Coupé (2005). However, the conditional fixed effects estimator performs quite well, exhibiting only a very slight drop in efficiency as compared to the random effects estimators. Finally, the penalized maximum likelihood estimator performs very well, clearly outperforming unconditional effects – suggesting that the penalized estimator is less affected by the incidental parameters bias – and performs as well or better than the random effects and conditional-FE estimators in mean-square errors terms. That is, even when the random effects assumptions

are known to hold with certainty both penalized maximum likelihood and conditional fixed effects perform comparably well to random effects despite the respective incidental parameters problem or sample losses.

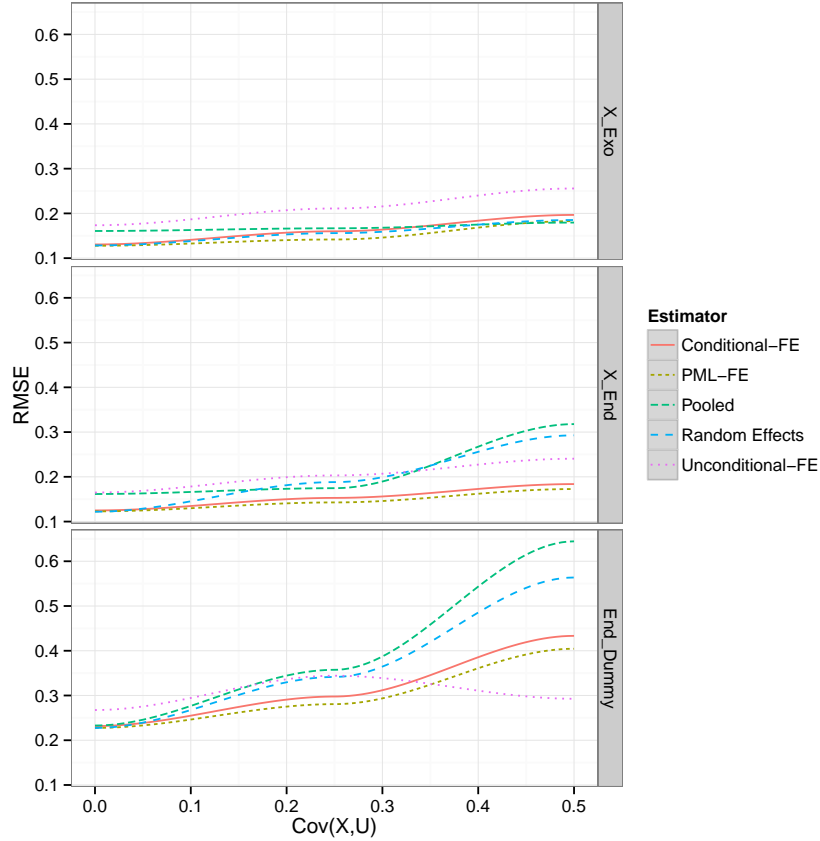
What happens when we relax the random effects assumption? In experiment's #2 and #3 – the mid and bottom row sets (Table 1) – some of the explanatory variables (X_2 and d) are made endogenous, setting ϕ to 0.25 and 0.5. In these simulations the dependent variable is invariant (e.g., always zero) for around seventeen and twenty-five units respectively, meaning with ϕ at 0.5 a full half of the sample is lost. As expected the endogeneity introduces bias into the random effects estimator, with ϕ at 0.25 already biasing the estimates of (β_2) and (β_4) by approximately 12% and 18% respectively.

Despite the sample truncation, the efficiency losses in the fixed effects estimators do not seem as considerable as prior work has suggested, instead the four-fold increase in the number of lost observations only causes the standard deviations of the sampling distribution to increase by about half for β_1 and β_2 (0.13 to 0.195 and 0.125 to 0.183). Surprisingly, the random effects estimator actually experiences a greater depreciation in its efficiency than conditional fixed effects. However, as the endogeneity (and sample truncation) increases, penalized maximum likelihood strictly dominates conditional fixed effects in mean square error terms, stemming largely from efficiency gains (presented in [Figure 5.1](#)). In all, the results suggest that in the presence of any non-trivial amount of endogeneity – i.e., non-orthogonality – between the unit effect and the regressors, those models which explicitly estimate unit effects are *far* more effective at recovering accurate parameter estimates.

We also wanted to test the extent to which the performance of these estimators is dependent on the unit effects being normally distributed. Though normality seems like a reasonable assumption to make about the distribution of unit effects, there are a number of alternative assumptions that are equally plausible. For instance, it may be the case that our data includes two or more distinct (normal) distributions which produce a non-normal mixture distribution. In the canonical example, the heights of men and women are both normally distributed within sex, but the height of humans is non-normal. Many political science examples are also distributed in this manner: developed vs. developing countries,

consolidated vs. emerging democracies, and, one might argue, states which experience conflict and those that do not. Therefore, we performed additional experiments where α_i was drawn from a $-\chi^2(3)$ (Figure 5.2) or bimodal ($\mu = 0, -6, p = 0.5$) distribution (Figure 5.3).

Figure 5.1: Accuracy of Coefficient Estimates ($\alpha_i \sim N$)



The results are consistent with, if not more pronounced evidence of (note the difference in scale between Figure 5.1, and Figure 5.3), the conclusions presented with the normal distribution. Though the pooled estimator performed poorly before, when we relax the assumptions of normality it performs considerably worse. For example, when the units are drawn from a mixture of two normals (e.g. bimodal), the pooled estimator underestimates the results of all β by near 60% when there is no correlation between the α and the regressors.¹⁹ When endogeneity is introduced the pooled estimator improves on the exogenous

¹⁹Not reported, tables made available in Appendix 1.

regressors, but the bias swings wildly in the other direction for the endogenous regressors producing estimates approximately 1.5 to 2 times(!) the truth for (β_2) and (β_4) . Random effects performs only slightly better under these conditions, overestimating the same regressors by approximately 51% and 62% respectively. The fixed effects estimators continue to do well, strictly dominating the random effects estimators in mean-square error terms even when the random effects assumptions hold (e.g. $\phi = 0$). Penalized maximum likelihood continues to do the best of all the alternatives, just bettering conditional logit due to its increased efficiency. However, the greater value of PML-FE over conditional fixed effects is not in terms of parameter estimates, but in that it allows us to calculate substantive marginal *effects*.

As has been noted throughout this text, we are ultimately interested in estimating meaningful substantive (ideally causal) effects. With binary outcomes this requires additional calculations, as coefficients do not directly equal effects as in the continuous-outcome linear regression model. While estimating these effects is always more onerous than with linear models, several of the estimators analyzed here are fundamentally incapable of providing these quantities of interest. Specifically, conditional fixed effects provides no estimate of for the average unit effect, meaning we are unable to estimate marginal effects of the regressors of interest and instead must instead rely on risk ratios (as discussed by [King \(2001\)](#)). This presents a problem for empirical researchers, the only consistent estimator available – when α is correlated with the regressors – is also the one which prevents them from calculating the quantities of interest we desire. Given the near parallel performance of penalized maximum likelihood with conditional fixed effects in parameters, we consider whether PML-FE may provide a solution to this problem.

Therefore, we estimate the marginal effects at the means for the 'truth' – e.g. with the parameters set to their actual values – and the estimates from each of the (available) estimators: pooled, random effects, unconditional fixed effects, and penalized maximum likelihood. The marginal effects for the continuous and dummy variable(s) are calculate respectively by:

Figure 5.2: Accuracy of Coefficient Estimates
 $(\alpha_i \sim \chi^2)$

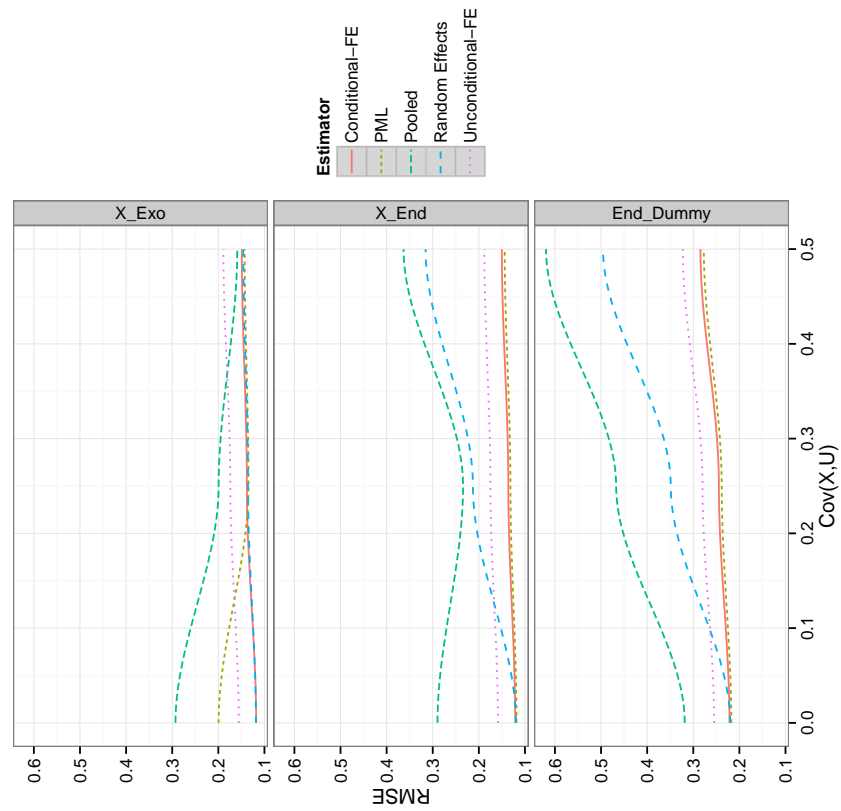
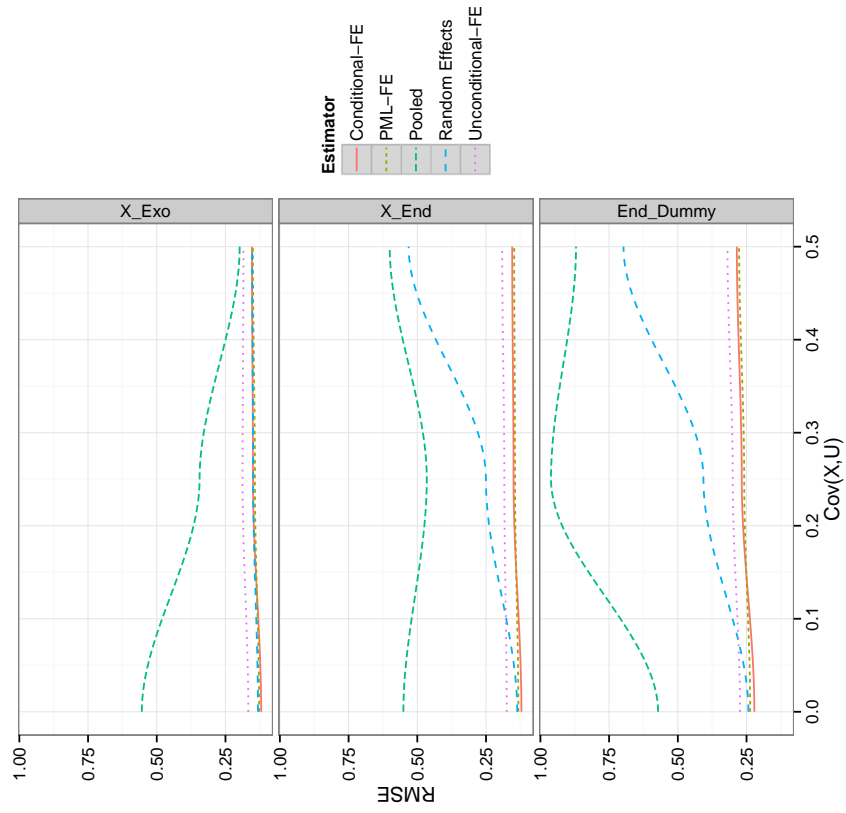


Figure 5.3: Accuracy of Coefficient Estimates
 $(\alpha_i \sim \text{Bimodal})$



$$\begin{aligned}\frac{\partial E[y_{it}|\mathbf{X}, d_{it}^4, \alpha_i]}{\partial x_{it}} &= \beta f(\Sigma\beta_k\bar{X}_k + \beta_4\bar{d} + \alpha) \\ \Delta E[y_{it}|\mathbf{X}, d_{it}^4, \alpha_i] &= F(\Sigma\beta_k\bar{X}_k + \beta_4 + \alpha) - F(\Sigma\beta_k\bar{X}_k + \alpha)\end{aligned}\tag{5.9}$$

The results are presented in mean-square error terms in [Figure 5.4](#), [5.5](#), and [5.6](#). The most prominent finding is the relatively strong performance of the random effects estimator in calculating accurate marginal effects, not only when the regressors are exogenous but often when they are endogenous as well. It consistently dominates the pooled estimates and frequently outperforms penalized maximum likelihood as well. The notable exception being dummy variables when endogeneity is high (0.5) and the distribution of the fixed effects is not normal, where penalized maximum likelihood provides a better option.²⁰ For all of the estimators we observe an upward bias on the effects estimates as endogeneity increases, suggesting the need for researchers to be cautious in the interpretation of these results regardless of the estimator they select.

How does random effects do so well in spite of its poor performance in estimating parameters? The result is largely a function of the accuracy of its estimate of the average unit effect, which is simply the constant for pooled and random effects and the sum of the individual unit effects for unconditional and penalized maximum likelihood fixed effects. The results are presented in [??](#). Random effects consistently outperforms the other estimators, producing estimates closer to the true average unit effect (-3 in our simulations). Unconditional effects easily does the worst (not reported) producing estimates of the average unit effect that are wildly biased and in the wrong direction (frequently returning results of positive 2 and 3). The reason for its poor performance is the sample truncation which results from rare-events, resulting in a sample with atypically high unit effects. This is a crucial distinction as previous work ([Greene 2004](#)) suggests the accuracy of conditional fixed effects, yet we show here that when the data are rare-events this is no longer the case. Penalized-maximum

²⁰This is exactly the type of regressor in dispute in the ‘Dirty Pool’ debates; the effect of a joint democracy dummy on peace. Furthermore, we suspect that part of the reason for this finding is that the endogenous dummy is ‘slowly changing’ and therefore may represent a broader class of covariates (not simply binary regressors). In [chapter 6](#) I explore the performance of these estimators on rarely-changing endogenous regressors.

Figure 5.4: Accuracy of MEMs
 $(\alpha_i \sim N)$

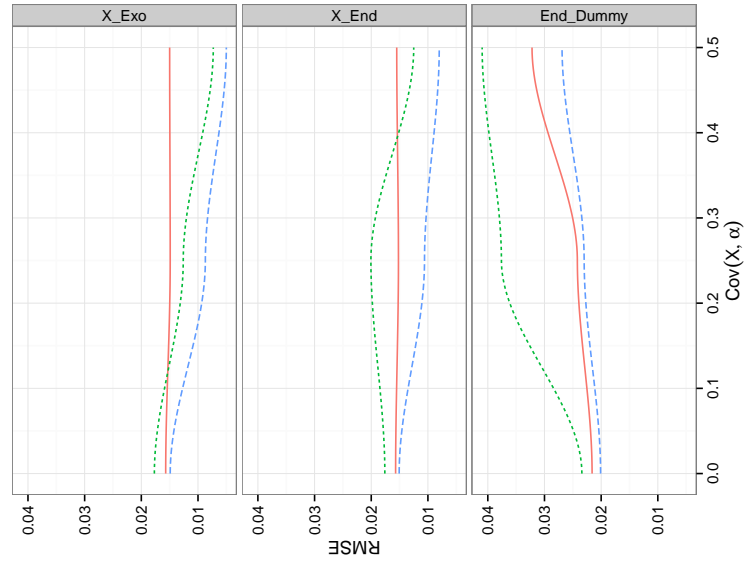


Figure 5.5: Accuracy of MEMs
 $(\alpha_i \sim \chi^2)$

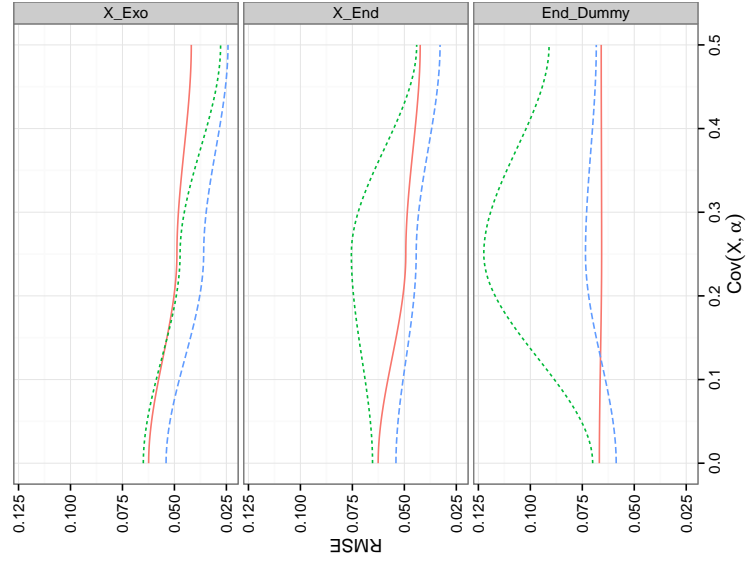
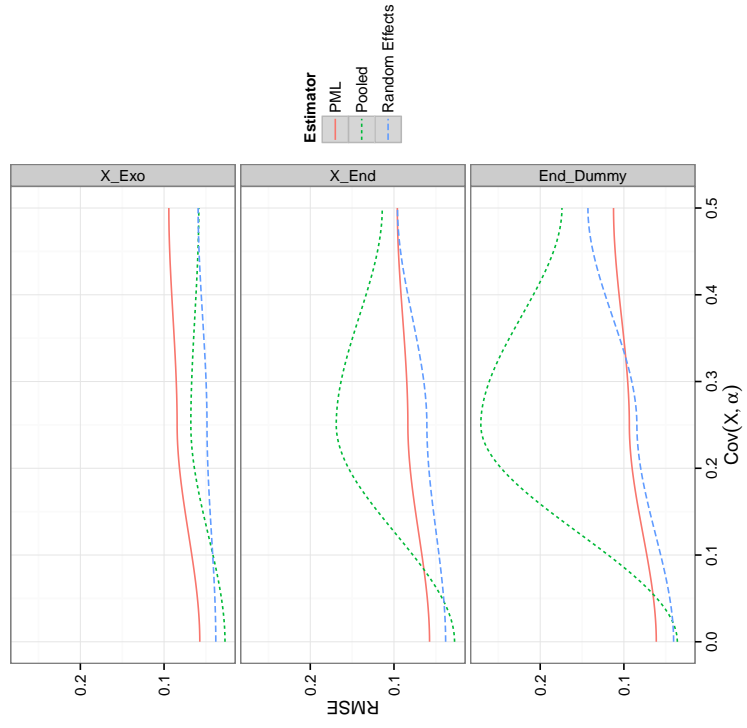


Figure 5.6: Accuracy of MEMs
 $(\alpha_i \sim \text{Bimodal})$



likelihood does reasonably well, but consistently seems to underestimate the average unit effect.

Importantly, the pooled and random effects estimators only provide a single unit effect estimate. In the former case, it is a common intercept, and in the latter, it is the mean of the distribution of random effects. By contrast, the PML estimator provides unit effect estimates for *each* unit in the sample, making it possible to calculate counterfactual effects for specific cases. This is a major potential advantage of the PML estimator that is simply not viable with any other estimator.

Table 5.3: Average Unit Effect Estimates

			Normal	Chi-sq	Bimodal
$\phi = 0$	Pooled	Avg. Unit Effect Est.	-2.65	-1.892	-1.204
		STD	(0.245)	(0.213)	(0.250)
	Random Effects	Avg. Unit Effect Est.	-3.039	-2.728	-3.204
		STD	(0.276)	(0.382)	(0.753)
	PML	Avg. Unit Effect Est.	-2.891	-2.532	-2.344
		STD	0.274	0.322	0.417
$\phi = 0.25$	Pooled	Avg. Unit Effect Est.	-2.619	-1.858	-1.181
		STD	(0.233)	(0.213)	(0.225)
	Random Effects	Avg. Unit Effect Est.	-2.962	-2.328	-2.323
		STD	(0.287)	(0.296)	(0.565)
	PML	Avg. Unit Effect Est.	-2.605	-2.054	-1.555
		STD	(0.219)	(0.234)	(0.280)
$\phi = 0.5$	Pooled	Avg. Unit Effect Est.	-2.313	-1.487	-0.625
		STD	(0.283)	(0.201)	(0.207)
	Random Effects	Avg. Unit Effect Est.	-2.623	-1.79	-0.871
		STD	(0.332)	(0.264)	(0.307)
	PML	Avg. Unit Effect Est.	-2.181	-1.548	-0.829
		STD	(0.224)	(0.235)	(0.265)

Note: True value for Avg. Unit Effects is -3.

5.3 DISCUSSION

Rare events are central to the study of International Relations, and political science more generally. Wars, revolts, coups, depressions, are all low probability-high impact events,

making them concurrently the most interesting but also the most difficult to explain, model, and predict. After discussing some of the problems raised in estimating these models, I proposed a novel strategy for estimating rare event-BTSCS data – penalized maximum likelihood fixed effects – which has attractive properties that should lead empirical researchers to privilege it over current ‘rare event’ and/or panel methods. While strategies currently exist to correct the small sample bias in maximum likelihood when estimating models of rare events (e.g., [King & Zeng \(2001b\)](#)), these do not naturally extend to cases where we believe unobserved unit heterogeneity is present. Penalized Maximum Likelihood, however, achieves comparable performance in reducing the small sample bias of ML and can easily model unit level heterogeneity through the inclusion of unit dummy variables.²¹ With the fixed effects specification, the penalization both reduces the incidental parameter bias and produces finite estimates event in instances of separation. Thereby enabling us to estimate individual unit effects (e.g., constants) for all units in the analysis regardless of whether they have switched state (e.g., experienced the event) during the temporal domain. Not only should assuage concerns about sample-selection, but also produce more efficient estimates of the regressors. The results from the preceding simulation suggest that this is indeed the case, as penalized maximum likelihood frequently performs as well or better than conditional fixed effects (a consistent estimator) despite the inclusion of incidental parameters.

Furthermore, the fact that penalized maximum likelihood produces unit effect estimates for each of the units in the sample has two potential additional benefits. First, it allows us to calculate unit-specific marginal effects for theoretically interesting cases rather than simply the arbitrary and ill-defined ‘average’ unit. It is the *only* panel estimator which can produce these values, which are often the most important for empirical researchers and policy makers. Second, using the unit effect estimates we can assess whether there is spatial autocorrelation in unobservables. This is a significant advancement, as current work in spatial econometrics offers no means of determining whether spatial clustering in the regressands reflects true contagion or clustering on time-invariant unobservable confounders, and is a fruitful area for

²¹In this respect, the penalized-pooled and penalized-panel estimators are nested, PML is the reduced model of PML-FE, facilitating straight-forward model discrimination (e.g., Likelihood Ratio test)

future research.²² However, both of these proposed extensions depend upon the accuracy of the estimates of the incidental parameters, which, to this point, remains an open question. As such, it will be important to analyze this explicitly in the future.²³

Lastly, given that both of the literature streams I respond to here were motivated by ‘democratic peace’ scholarship – e.g., rare events with dyad-year as the unit and resultantly large sampling dimensions – it seems natural to re-analyze that question using the penalized maximum likelihood strategies presented here. However, two limitations constrain my ability to do so at present, one computational and one econometric. First, available programs to estimate these models cannot currently estimate models with as many parameters as would be required for such an analysis, which can include upwards of 10,000 unit dummies alone.²⁴ Second, while the evidence here suggests that penalized likelihood helps to ameliorate some of the ills of the incidental parameter bias, it can still present problems in small t samples. These problems may be more salient in the sample dimensions common to dyadic analysis, where the N to T ratio is much more pronounced than the sampling dimensions analyzed here. This later issue is less a concern than an open question, and one that should be explored when available technologies permit. However, an easy evasion that solves addresses both concerns by simply reducing the number of incidental parameters to be estimated: group fixed-effects.

When estimating (B)TSCS models researchers have traditionally been forced to defend one of two rather extreme positions regarding unit effects: *i*) complete homogeneity (e.g., pooled) *ii*) complete heterogeneity (e.g., fixed effects).²⁵ Is this always reasonable? As I have noted throughout, ultimately the decision of which approach to prefer is ultimately a theoretical choice motivated by ones understanding of the data. It seems likely that for a

²²This is similar to how spatial-Durbin models currently allow us to discriminate between spatial correlation in the outcomes and in the regressors.

²³All the data to do so exist now, I just need to come up with an efficient means of comparing the estimates for the each of the parameters to the true fixed effects specified in the data generating process.

²⁴I have been discussion with the creator of the R package “brglm” – the package used to estimate penalized maximum likelihood – about modifying the estimator to facilitate such procedures and he has stated that he plans to incorporate sparse-matrix capabilities into the package soon which would facilitate the estimation of such models.

²⁵There are notable ad-hoc exceptions to this in applied work, such as estimating regional fixed effects, however to my knowledge it has not been systematically explored as a general practice in political science.

number of applications, neither complete homogeneity or heterogeneity is entirely accurate, but instead that the units can be divided sub-groups which are comparable to one another but distinct from the remaining groups (units, etc...). We can represent this as a generalized version of Equation 1.1, given by:

$$y_i^* = \alpha_g + \mathbf{X}_i\boldsymbol{\beta} + \epsilon_i \quad (5.10)$$

where subscript- g identifies a group from the complete set of groups G . Note the generality of this set up, $G = 1$ groups produces the pooled model, while $G = N$ produces the fixed effect model. When $1 < G < N$ there are group-specific patterns of heterogeneity (e.g., common preferences, utilities, risk propensities) which we capture with a common-group intercept.²⁶ As such, this represents a compromise approach between the more traditional extremes.

As noted analytically by [Bester & Hansen \(2013\)](#), group-fixed effects represents a trade-off between two types of bias. With too many groups (at the limit unit-fixed effects) our model suffers from the incidental parameter bias, with too few (at the limit pooled) there is omitted variable bias from the unmodeled unit heterogeneity. Intuitively then, we want to minimize total bias by selecting a group scheme requiring as few parameters as necessary to adequately capture the heterogeneity across units. While researchers could identify these groups in an ad hoc manner (e.g., regional dummies), ideally we would prefer a strategy which helps locate group clusters amongst the data. [Bonhomme & Manresa \(2012\)](#) have recently suggested a strategy to estimate group membership that minimizes a least-squares criterion with respect to all possible groupings of the units. However, at present, such strategies have only been elaborate for linear models. As such, extending an approach such as theirs to allow for the estimation of non-linear models – or employing some other technique to identify latent group-clusters in the data such as Lasso – may be worthwhile.

²⁶In some ways the theoretical motivation for such an approach is quite similar to the previous discussion of spatial interdependence, generally, that cross-sectional units have underlying patterns of dependence. However, here our assumption is that units cluster on unobservables (e.g., common shocks) which can be captured in full with the inclusion of a group dummy.

6.0 THE KNOWN UNKNOWNNS OF CIVIL WAR

The failure to dissect the cause of war leaves us open for the next installment.

— Chris Hedges, 1970

There are known knowns, there are the things we know we know, and we also know there are known unknowns, that is to say, we know there are some things we do not know. But there are also unknown unknowns, the ones we don't know we don't know.

— United States Secretary of Defense, Donald Rumsfeld, 2002

The centrality of income in the literature on the causes of civil war is clear. Income per capita is widely seen as *the* key factor in determining where conflict is likely to occur, with scholars arguing that it is the “most important variable” from a theoretical perspective and empirically the most robust (Hegre & Sambanis 2006). In both Fearon & Laitin (2003) and Collier & Hoeffler (2004) – the most widely cited empirical work on civil war – income per capita is argued to play a central, if different, role in the production of conflict. In short, civil war is widely regarded as a “problem of the poor” (Sambanis 2002, 216). From this, researchers and policy makers alike have concluded that “the key root *cause* of conflict is the failure of economic development” (Collier *et al.* 2003, 53). Yet, how much support do we have for a causal theory of development and civil war? I argue that in many respects this relationship has gone presumed rather than proven. While to some the link between development and conflict may “seem obvious,” because “if you read the newspapers, you will

see that the countries where there is conflict are far more likely to be poor,” the importance of this relationship calls for closer scrutiny (Collier 2008, 18)

That the preponderance of civil wars occur in poverty-stricken countries is undeniably true. Civil war occurs almost 10 times as frequently in the worlds poorest countries as it does in its richest (Fearon 2008).¹ Sierra Leone, the Democratic Republic of Congo, Indonesia, the Ivory Coast, Liberia, and far too many more countries have known both crippling poverty and destructive internal war. However, the mere observation that those countries which are low in development are high in civil war does not necessarily suggest a direct relationship between the two, quite simply, correlation does not imply causation. In what follows, I review the literature on income and civil war and find surprisingly little support for a direct relationship in canonical bargaining or contest models of rebellion (Chassang & Padro-i Miquel 2009, Fearon 2008). Rather, I argue that the observed relationship between the two is spurious, with other latent conflict-resolution technologies determining both the level of development and the propensity of conflict. As such, in most current analyses, income per capita simply proxies for unobservable (and therefore unmodeled) determinants of conflict – e.g., malfunctioning social institutions, inter-ethnic tensions, historical animosity – which are distributed in a similar pattern.

While most civil war scholars widely admit to the likely presence of such unobservables – with some theories directly suggesting their influence – substantially less attention is paid to handling these unobservable factors in our empirical analyses. This is despite the fact that a common strategy exists to control for these unobservables and thereby mitigate their ill-effects, namely, country fixed effects. The use of such approaches has become increasingly widespread in other fields (see notably Acemoglu *et al.* 2008), yet civil war scholars have, by and large, failed to similarly embrace these methods.² In part, the hesitancy of conflict researchers stems from important concerns raised about the permissiveness of fixed effects when dealing with rare-event binary data (Beck & Katz 2001, Green *et al.* 2001, King

¹As determined by comparing the frequency of conflict in the one-fifth poorest country-years to the one-fifth richest country-years.

²There are, to be sure, notable exceptions to this. However, the point remains that as a general practice fixed effects models are still not the norm despite the advantages they should offer given the widely-held assumption of unobservable determinants of civil war.

2001, [Oneal & Russett 2001](#)) and rarely-changing regressors ([Beck & Katz 2001](#), [Plümper & Troeger 2007](#)), both of which are present here. If correct, these arguments would suggest that the inability of researchers to find a relationship after including country fixed effects in these models is not meaningful theoretically, but simply a statistical artifact of the data and approach. As such, it is unclear what researchers should make of the finding that the effect of GDP on civil war goes away when fixed effects models are specified [Djankov & Reynal-Querol \(2010\)](#). By itself this finding means little, as civil war scholars have long been aware that this is the result from fixed effects modeling. The more central question, in light of the previous methodological work mentioned above, is what to make of this finding: does it say something theoretically meaningful or is it an artifact of an unsuitable modeling strategy?

As such, I elaborate and extend on the main issues raised in estimating fixed effects models of rare-events binary time-series-cross-sectional (re-BTSCS) data with nearly time-invariant regressors. After which, I directly engage the more serious of these concerns. First, utilizing the strategy discussed in [chapter 5](#), I discuss how a penalized maximum likelihood fixed effects model *i*) obtains first-order unbiased estimates in small samples (e.g., rare events), *ii*) allows for the identification and estimation of unit-specific intercepts for even those units which are time-invariant in the outcome, and *iii*) permits the estimation of case-specific counter-factual substantive effects. Second, I explore the behavior of common BTSCS estimators when dealing with endogenous and slowly-changing variables; those with significantly greater between variation than within variation. In the presence of such variables, it has been argued that fixed effects estimation may perform poorly [Plümper & Troeger \(2007\)](#). However, as noted by [Beck \(2011\)](#), these issues are still new and require further attention. Therefore, I offer some practical guidelines on the conditions under which such variables are likely to complicate inference with non-linear models. Subsequently, I bring these insights to bear on the question of the relationship between income and civil war, replicating the analyses of [Fearon & Laitin \(2003\)](#) and interpreting the results in light of the prior discussion. Anticipating my findings, I conclude that any direct relationship previously found between development and conflict is likely spurious.

This has important policy implications as the promotion of development has been seen as the linchpin in reducing civil war. In part, this has been the result of recommendations from scholars who have repeatedly and strongly attested to this connection. As noted above, [Collier *et al.* \(2003\)](#) in a report to the World Bank states that the causal relationship between the development and civil war is their key argument. Similarly, in their paper “What Policy Makers Need to Know,” [Rice *et al.* \(2006\)](#) argue that the debate on the principal causes of civil conflict has been resolved: poverty matters. Furthermore, in their policy implications they go on to suggest that there is “little doubt that policies that increase per capita income in the poorest countries will reduce their conflict risk” (13). This despite the fact that we have no evidence suggesting the efficacy of such an intervention, as there is no effect of *within* country variation in GDP on civil war. That being said, this should not be taken as advice to scale back efforts at economic development. In addition to the many other benefits that a country is known to derive from greater development, there may also be an indirect relationship with civil war. To that end, I conclude by sketching a possible conditional theory of income and civil war.

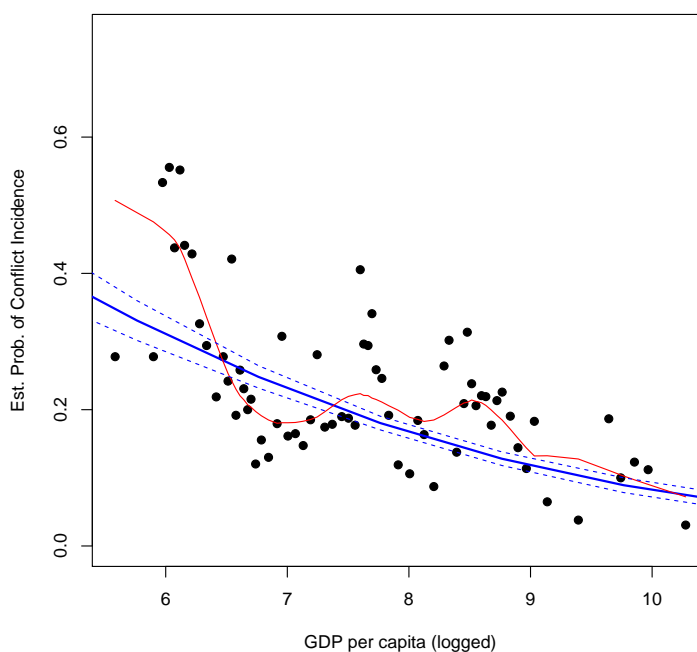
6.1 CUM HOC ERGO PROPTER HOC: LOW INCOME AND CIVIL WAR?

That civil wars frequently occur in poor countries is itself an indisputable fact. The poorest countries in Africa, Asia, and Latin America are also those disproportionately more likely to experience civil war. [Figure 6.1](#) indicates the strong association between GDP per capita and the probability of civil war onset.³ Moreover, this pattern has been supported by a range of cross-national empirical studies which has concluded broad support for the positive (negative) relationship between poverty (wealth) and civil war ([Collier & Hoeffler 2004](#), [Elbadawi & Sambanis 2002](#), [Fearon & Laitin 2003](#), [Hegre & Sambanis 2006](#), [Hegre 2001](#), [Thies 2010](#)). That is, controlling for other *observable* factors believed to produce conflict, the relationship between development and civil war remains. [Hegre & Sambanis \(2006\)](#) conclude

³The data in the figure are the same used in the subsequent estimation.

that GDP per capita as the most robust indicator of civil war.⁴ Collier *et al.* (2003) go even further suggesting that “all studies agree that a link exists between poverty and civil war” (58). Yet, even among this seemingly complementary work, there remains uncertainty and disagreement on the principal mechanism underlying this relationship.

Figure 6.1: Development and Civil War Incidence, 1950-2008



Broadly, two main explanations have been advanced linking development to civil war.⁵ First, Collier & Hoeffler (2004) advocate an economic explanation, wherein conflicts are argued to be more likely in low income countries as a result of lower opportunity costs for would-be combatants. That is, it is easier for rebel organizations to recruit new members in low income countries because there is less benefit to remaining in the labor market (or forgone by leaving it).⁶ Alternatively, Fearon & Laitin (2003) have suggested that the relationship between GDP and civil war more likely reflects variation in state capacity. Better developed

⁴Hegre & Sambanis (2006) find population to be nearly as robust but less substantively meaningful.

⁵Holtermann (2012) offers a recent summary of these different perspectives.

⁶While this should make it easier for both sides – rebels and government – to recruit, (Collier 2000) argue that asymmetries in initial capacity mean that cheap recruits are more significant for rebels than governments Holtermann (2012).

states have greater organization, reach, and technologies with which to locate and disrupt potential rebel organization. Thus, in states lacking these capabilities, rebels should be better able to mobilize resources and support to help fuel the insurgency effort. In sum, the two major works in civil war have produced a similar finding but have taken from it fundamentally different conclusions.⁷ Despite the centrality of this issue for understanding civil war, efforts to resolve this debate and illuminate the relationship between development and conflict have been surprisingly rare (Fjelde & De Soysa 2009, Holtermann 2012, Thies 2010). As such, we are left with an apparent empirical reality searching for an explanation.

The inability of researchers to pin down a clear explanation linking development to war may suggest the need to exercise greater caution in asserting the strength of relationship. Justino (2006) argues that this research “offer(s) only limited systematic accounts of the mechanism through which low incomes amongst a large fraction of society affect the outbreak of war” and are based on assumptions which are “largely untested” (25). Moreover, formal analysis into civil war, including that by Fearon himself, should give researchers even greater pause, as both bargaining and contest models of rebellion have rejected the idea of a direct link between per capita income and conflict. Building on the canonical bargaining model of war (Fearon 1995, Powell 2006), Chassang & Padro-i Miquel (2009) conclude that conflict is not a function of the productive capacity of the state. In short, while the opportunity cost of rebellion is diminished in poor states so too are the benefits of success. In this respect the “costs and benefits from fighting move proportionately to the size of the economy, yielding no natural link” (Chassang & Padro-i Miquel 2009, 220). To better understand this intuition consider their simple static model of conflict bargaining.⁸

Assume two groups $i \in \{1, 2\}$ each sharing a territory of size 2, where group 1 controls $1 + \lambda$ parcels of land and group 2 the remaining $1 - \lambda$ (with $\lambda \in \{0, 1\}$). The land is used to produce crops (though any substitute could suffice), which are generated according to the production function $C(\theta, L, l) = \theta Ll$ where L is the amount of land, l is the amount

⁷Even this neglects the possibility of a ‘grievance’ motive for potential rebellion, wherein individuals in low-income countries are more likely to rebel because of more salient political, social, and economic hardships (?). It is this line of research which Collier & Hoeffler (2004) were originally responding to, arguing that as grievances are ubiquitous they are poor explanators for war.

⁸Chassang & Padro-i Miquel (2009) extend this to the fully dynamic case, but the static model is sufficient to show the intuition for why conflict is unrelated to income.

of labor, and θ is the fertility of the land (more generally, the size of the economy). Each group controls 1 unit of labor s.t. if all labor is used in production 2θ is realized. However, either group may choose to forgo production and instead divert some $c \in \{0, 1\}$ amount of their labor toward an effort to seize land from the other group with a $P > 0.5$ probability of success (assumes a first mover advantage). Therefore, the opportunity cost of fighting is $2C\theta$ (e.g., the forgone production). As such, opting for war yields a payoff of $2\theta(1 - c)$ with probability P and 0 with a probability $1 - P$.

As in [Fearon \(1995\)](#) conflict can be avoided through bargaining wherein one group makes a transfer of land T to the other such that both groups are made better off in expectation than they would be from going to war. If no such transfer exists, then conflict occurs. Therefore, for peace to be maintained the following two conditions must hold:

$$\textbf{Group 1:} \quad (1 + \lambda)\theta - T\theta > P2\theta(1 - c) \quad (6.1)$$

$$\textbf{Group 2:} \quad (1 + \lambda)\theta + T\theta > P2\theta(1 - c) \quad (6.2)$$

That is, the transfer T must be large enough to dissuade group 2 from preferring war, but small enough to preclude group 1 from doing the same. Both conditions are true if and only if:

$$\theta > P2\theta(1 - c) \quad (6.3)$$

The implication is immediately evident, θ (e.g., the size of the economy) does not influence the likelihood of fighting as the opportunity cost of fighting and the spoils from victory are linearly related to one another. As noted by [Chassang & Padro-i Miquel \(2009\)](#), both the spoils (2θ) and cost ($2c\theta$) are increasing functions in the size of the economy, meaning no change in the size of the economy alters in the inequality given in condition (6.3).⁹ As such, cross-national variation – that is, between variation – in states conflict

⁹Instead, [Chassang & Padro-i Miquel \(2009\)](#) continue that fighting obtains when the following condition holds:

propensities *cannot* be explained directly by the level of development. Instead, [Chassang & Padro-i Miquel \(2009\)](#) conclude greater support for within country theory of income, namely, economic shocks. Analogous to [Powell \(2006\)](#)’s more general model on shifts in power, large short-term fluctuations in productivity (and sufficiently discounted future returns) can induce actors to opt for conflict. This may suggest researchers should prefer strategies which discount between variation in favor of better isolating within variation when attempting to identify the determinants of conflict.

[Fearon \(2008\)](#) reaches a similar conclusion when model conflict as a contest model ([Grossman 1991](#), [Hirshleifer 1995](#), [Skaperdas 1992](#)). As with [Chassang & Padro-i Miquel \(2009\)](#), he finds that given that the realized gains from fighting increase in proportion to the wealth of the state, there is no reason to suspect reduced violence (e.g., the bigger the pie the greater incentive to fight). Furthermore, he shows that even incorporating marginal utility of income understandings into our risk probabilities doesn’t completely solve this issue.¹⁰ As a result, [Fearon \(2008\)](#) argues that this result undermines support for poverty based explanations of war commonly given in the empirical literature. Instead he offers possible second-order explanations, that is, the pacific effects of characteristics associated with high levels of GDP but which do not directly result from it.

In sum, neither the canonical bargaining nor contest model of civil war can provide support for a direct relationship between income and civil war. As such, it seems more likely that the observed empirical relationship between the two is not causal, as is frequently argued, but emerges because the same underlying factors which give rise to conflict also impeded development. As argued by [Acemoglu *et al.* \(2008\)](#) when discussing the relation between democracy and income, historical country-specific factors which make some countries more likely to experience both positive (or negative) outcomes often go unmodeled as they are

$$P > P^S \equiv \frac{1}{2(1-c)}$$

That is, conflict is determined by the relationship between the first-mover advantage and the opportunity cost.

¹⁰Though ultimately, if we make the assumption that poor people are *relatively* more risk averse it would introduce a non-linearity into the utility function and GDP would re-emerge as a predictor. [Fearon \(2008\)](#), however, considers such an approach and finds there to be little consistent justification for it.

difficult to observe and typically extend well beyond conventional sample dimensions. In the case of civil war and development, I argue that some states, over the course of history, developed better conflict-management processes.¹¹ The ability to resolve low level disputes without resorting to violence aided development and reduced the risk of future fighting. That is, to the extent that there is a relationship between development and fighting it is one borne out of hundreds of years. Conversely, those states which were unable to resolve such disputes peaceably were set back in their development – forced to devote resources to security, dispute resolution, fighting, etc... – and had a greater probability of future conflict. As such, I argue the empirical relationship between GDP and civil war noted by conflict scholars is spurious. Once the unobserved country-specific factors likely to influence are controlled for (via fixed effects), I expect there to be no direct relationship between the two.¹²

6.2 RARE EVENTS AND RARELY CHANGING REGRESSORS

Despite the apparent importance of unobservables in driving conflict processes, there has yet to emerge a consistent strategy for addressing them. Numerous papers make no attempt to model unit effects at all – preferring pooled logit or probit estimation – while those that do adopt quite different approaches. For example, Fearon and Laitin (2003) re-estimate their main model(s) using conditional fixed effects logit and indicate that their results are ‘virtually identical’ to the pooled estimates.¹³ Instead, Sambanis (2001) re-estimates his models using random effects probit, ultimately preferring the simple probit estimator – despite rejecting the null of independence – because of the similarity in the results. Finally, Collier and Hoeffler (2004) argue that fixed effects estimation is ‘very severe’ in their interpretation of these results. Thus, even the canonical works in the civil war

¹¹These processes ultimately took the form of institutions, which also explains [Acemoglu et al. \(2008\)](#)’s finding on democracy

¹²While frequently tests of ‘no significance’ would be odd, as there are lots of reasons one may fail to find a relationship, in this instance the strength of the finding in the conventional literature makes a non-finding interesting in its own right.

¹³However, they do not report these findings in text.

literature – with each of these articles having been cited more than 500 times – disagree over the role of unobservables in the determination of civil war and how they should be addressed in empirical research. Furthermore, researchers frequently seem to confuse the issues underlying the decision over which modeling strategy to prefer.

In part this confusion stems from a lack of consensus among methodologists on how to handle these issues. In the ‘Dirty Pool’ symposium more than a decade ago, the topic of fixed effects models for re-BTSCS data was debated (Beck & Katz 2001, Green *et al.* 2001, King 2001, Oneal & Russett 2001). Through the course of these discussions a number of potential issues for fixed effects estimation were raised, with Beck & Katz (2001) concluding that it is *never* a good idea to estimate fixed-effects models with rare-event binary time-series cross-sectional data. It is easy to see why such strong claims would have lasting effects on civil war scholars. In particular, two of the problems discussed seem to have resonated widely and are still voiced as concerns against estimate fixed effects models.

First, those units which do not experience a civil war are dropped from the analysis, possessing no within-variation in the outcome.¹⁴ While some suggested at the time this was not a problem Green *et al.* (2001), others argued that dummy variables are atheoretical and removed all the between-unit variation from the model.¹⁵ Many in the civil war literature to raise this issue as their reason for avoid fixed effects, including Collier & Hoeffler (2004). Others refer back to the critique of Beck & Katz (2001) explicitly, such as Nel & Righarts (2008) who argue “we do not run fixed effects models. Following Beck & Katz (2001), we consider the use of fixed effects models to control for the influences of unit idiosyncrasies in binary-outcome time-series cross-sectional data as pernicious. There are many units with no other outcome than zero, and to control for their presumed effects on the parameter estimates does not make any sense.” Similarly, Buhaug & Gleditsch (2008), citing Beck &

¹⁴While Heckman (1981) notes this ceases to be a problem as $T \rightarrow \infty$, given that we are always estimating models with finite- T it may be important for empirical research

¹⁵Specifically, in response to Green *et al.* (2001), Beck & Katz (2001) defended the democratic peace writing, “Green, Kim and Yoon...argue that if we discovered new democratic dyads that were always pacific, it would give us no information, because ‘we do not know the base probability (the intercept) of war for each of these new dyads.’ We freely admit that it is logically possible that these new dyads might be pacific because of the name of the dyad (the fixed effects) or because both partners grow green beans. But it seems odd to throw out the only theoretical explanation we have, that the dyad is pacific because it is democratic” (490).

Katz (2001) note that they are “generally skeptical of such methods,” as it “requires us to treat as non-informative all countries where we do not observed variation in the response” (227).¹⁶

The second limitation of fixed effects which is consistently raised is the inability to include time-invariant or nearly time-invariant regressors (Beck & Katz 2001, Plümper & Troeger 2007). While the former is certainly true, the latter is potentially more interesting, as FE estimators do provide estimates for rarely changing regressors. However, as noted by Beck (2001) “Although we can estimate a model with slowly changing independent variables, the fixed effect will soak up most of the explanatory power...[and] make it hard for such variables to appear either substantively or statistically significant” (285). That is, by eliminating the between variation and retaining only the less substantial within variation, fixed effects “masks the impact of slowly changing independent variables” (Beck 2001). Given that a number of the most significant determinants of civil war change slowly over time (e.g., institutions, development, population...), this is a salient issue for conflict studies. As with the issue of sample selection, researchers use these concerns not just to justify alternative estimation strategies, but even where fixed effects models are run they voice them in an apparent effort to explain away aberrant findings.¹⁷

As noted, civil war data are both rare and models of conflict typically include rarely changing variables. Thus, to the extent that either of these issues truly renders fixed effects problematic, the hesitancy on the part of researchers is well placed. Having said that, how important are these issues for fixed-effects models? The issue of sample selection is largely taken up in chapter 5, so I only briefly touch on it here. First, to the extent that sample selection is producing the incongruous results, there should be another way to prove this. Namely, estimating a pooled model including only those units which experience a conflict during the sample (e.g., the same selection step which occurs when we estimate a fixed-effects model). If the findings remain with this conflict-only sub-sample then it simply cannot be

¹⁶I should not that following this discussion Buhaug & Gleditsch (2008) does indeed estimate a fixed effects model, despite their misgivings.

¹⁷Collier *et al.* (2009) seems to be an example of this, following their fixed effects model they note that “none of the variables which change slowly over time are significant,” but do appear to take this inability to produce results as meaningful.

the case that selection alone is distinguishing the two results. Second, even with severe sample selection, (conditional) fixed-effects estimators produce unbiased results. There is some loss in efficiency, as it does not use the entire sample, but they are still substantially better off in mean square error terms when unobserved heterogeneity is present. Second, if we consider the absence of conflict in particular states an artifact of the temporal domain – e.g., on a long enough time horizon all units would experience failure – then strategies exist to model this belief and estimate fixed effects while retain the entire sample. This penalized maximum likelihood fixed effects (PML-FE) is discussed in detail in [chapter 5](#) so I do not discuss it further here.

The issue of rarely-changing variables has been less widely discussed in the literature. While this should not be cause for determining which model to estimate, it is important for understanding how to interpret the results we obtain. If, for example, it is the case that fixed effects masks rarely changing variables, then we would want to discount the importance of particular variables losing significance. [Plümper & Troeger \(2007\)](#) provide the most substantial – and possibly only – systematic analysis into the consequences of fixed effects estimation for rarely changing variables. Their analysis focused primarily on comparing fixed effects to their fixed-effects variance decomposition estimator for continuous-outcome data.¹⁸ In general, [Plümper & Troeger \(2007\)](#) find that the performance of fixed effects diminishes as a variable becomes increasingly “slow” but improves as it becomes increasingly endogenous. While I fully expect these same themes will generally hold, a number of outstanding issues require further inquiry. First, what impact, if any, does sample selection (e.g., rare events) have on the performance of these estimators? Second, and more importantly, what is the false-positive rate for the non-FE estimators when unit heterogeneity goes unmodeled? That is, how frequently will we wrongly conclude significance for a irrelevant (endogenous) variable when we fail to account for unobserved heterogeneity?¹⁹

¹⁸The estimates for pooled and random-effects estimation are included in the online appendix, yet presented in such a highly aggregated form that it is difficult to infer much beyond the general themes I note in text.

¹⁹An additional concern, but one that is not addressed here, is the accuracy of these estimators in calculating the substantive effects of slowly changing regressors. In future work, I explore this question in greater detail.

To answer these questions and provide some general guidelines for researchers when dealing with rarely changing regressors and non-linear outcomes, I estimate a variety of monte carlo simulations. The basic framework for these experiments is given in [chapter 5](#), with a slightly different specification here to explicitly assess the effect of rarely changing regressors. The DGP includes a time-varying exogenous regressor (X_{it}^1), a time-varying endogenous regressor (X_{it}^2), a nearly time-invariant exogenous regressor (X_{it}^3), a nearly time-invariant endogenous regressor (X_{it}^4) and a fixed unit effect α_i . More specifically, the DGP is:

$$y_{it} = \mathbf{1}[\beta_1 X_{it}^1 + \beta_2 X_{it}^2 + \beta_3 X_{it}^3 + \beta_4 X_{it}^4 + \alpha_i + \epsilon_{it} > 0] \quad (6.4)$$

Where $i = 1, \dots, N$, $t = 1, \dots, T$ and...

$$\begin{aligned} X_{it}^1 &\sim N(0, 1) \\ X_{it}^2 &= \tilde{X}_{it}^2 + \phi\alpha_i, \text{ where } \tilde{X}_{it}^2 \sim N(0, 1) \\ X_{it}^3 &= \tilde{X}_i^3 + \tilde{X}_{it}^3, \text{ where } \tilde{X}_i^3 \sim N(0, 1) \text{ and } \tilde{X}_{it}^3 \sim N(0, \sigma) \\ X_{it}^4 &= \tilde{X}_i^4 + \tilde{X}_{it}^4 + \phi\alpha_i, \text{ where } \tilde{X}_{it}^4 \sim N(0, 1) \text{ and } \tilde{X}_{it}^4 \sim N(0, \sigma) \end{aligned}$$

as indicated X_{it}^3 and X_{it}^4 are the ‘rarely’ changing measures. In all experiments the between variation is held fixed at 1, while the within variation σ is modified $\{1.0, 0.8, 0.6, 0.4, 0.2\}$ to capture different levels of near invariance, with lower values indicating a more slowly-changing measure.²⁰ As before, the fixed-effect (α_i) is varied across the experiments, drawn from from one of two following distributions:

$$\begin{aligned} \alpha_i &\sim N(-3, 1) \\ \text{or...} \\ f(\alpha_i) &= p(g_1(\alpha)) + (1 - p)(g_2(\alpha)) \\ \text{where } g_1(\alpha_i) &\sim N(0, 1), g_2(\alpha_i) \sim N(-6, 1), \text{ and } p = 0.5 \end{aligned}$$

That is, the performance of the estimators is evaluated when the unit effects are distributed normal or bimodal (with each peak being normally distributed). As noted in [chapter 5](#), the

²⁰This is similar in nature to the approach taken by [Plümper & Troeger \(2007\)](#), that is, modifying the ratio of between to within variation.

random effects estimator assumes a distribution of the unit heterogeneity (e.g., normality), when this is wrong it will negatively affect these estimates. In applied work, of course, we never know the distribution of these effects. As such, it is useful to test these models under a range of possibilities to examine their performance. For civil war a bimodal distribution seems just as, if not more, reasonable theoretically than assuming normality, with some states being ‘high risk’ and other states at effectively no risk at all.²¹ Unless otherwise noted, in the reported experiments the level of endogeneity ϕ is 0.25 and the sampling dimensions are $N = 100$ and $T = 20$.

Table 6.1 reports the estimates of the nearly time-invariant regressors (β_3 and β_4) for the 1st set of these experiments, which assume the unit effects are distributed normally. We see that all of the estimators perform worse when the within-to-between variation is low (e.g., more slowly moving), as presented in the leftmost column.²² While the pooled and random effects estimators perform the best in mean-square error terms, both evidence a significant upward bias in the estimate of the endogenous regressor. Note the implication of this, in the conditions where the current literature would most strongly advise researchers prefer pooled or random effects (to fixed effects) these estimators are at the *greatest* risk of committing type-I errors (i.e., false positives). Both the pooled and random effects estimates improve as the within-to-between variation levels (columns further right). In particular, the improvements of the pooled estimator are interesting. For the exogenous regressor β_3 it continues suffers from the expected attenuation bias at all levels of ‘slow,’ with its improved performance resulting from efficiency gains alone.²³ However, the estimates of the endogenous regressor β_4 improve considerably (become less biased) as the within variation rises in proportion to the between variation. This may seem odd, but what it is actually capturing is that the relative strength of the endogeneity is diminishing as the within variation rises and, as a result, the pooled estimator performs better.²⁴ As a test of this I hold the within-

²¹As noted in chapter 3 this is exactly what is argued by Collier *et al.* (2003)

²²Occasionally, I may invert this and note the between-to-within variation as being high for slowly-changing regressors, meaning the same thing of course.

²³As discussed in chapter 5 this the downward bias induced from omitting exogenous incidental parameters as discussed by Wooldridge (2010) and others.

²⁴In the future different specifications for the slow-moving variables will need to be explored. In particular, evaluating the performance of the estimator when a variable is slowly changing do to persistent dynamics in the regressor.

to-between variation fixed and increase the correlation ϕ between the fixed effect α_i and the endogenous X_4 . These results indicate that this is indeed the case, as the performance of the pooled estimator degrades noticeably, evidencing the expected inflationary bias.²⁵

Table 6.1: Panel Estimators with Rarely Changing Regressors ($\alpha_i \sim Normal$)

$\phi = 0.25$		$\sigma_{within}/\sigma_{between}$:	0.2	0.4	0.6	0.8	1.0
Pooled	β_3	Coeff. Est	0.910	0.915	0.897	0.908	0.908
		Standard Deviation	0.168	0.151	0.128	0.116	0.099
		RMSE	0.191	0.174	0.164	0.148	0.136
		SE	0.115	0.109	0.100	0.091	0.084
	β_4	Coeff. Est	1.114	1.089	1.055	1.023	1.003
		Standard Deviation	0.167	0.152	0.130	0.118	0.099
		RMSE	0.202	0.176	0.141	0.120	0.099
		SE	0.118	0.111	0.102	0.094	0.086
	Random Effects	Coeff. Est	0.985	0.994	0.988	1.001	1.002
		Standard Deviation	0.173	0.156	0.133	0.126	0.105
		RMSE	0.174	0.156	0.133	0.126	0.105
		SE	0.170	0.152	0.134	0.118	0.105
Con-FE	β_3	Coeff. Est	1.201	1.164	1.124	1.089	1.075
		Standard Deviation	0.176	0.159	0.136	0.122	0.106
		RMSE	0.267	0.228	0.184	0.151	0.130
		SE	0.173	0.154	0.135	0.118	0.106
	β_4	Coeff. Est	1.007	1.004	1.024	1.013	1.007
		Standard Deviation	0.604	0.294	0.201	0.161	0.125
		RMSE	0.604	0.294	0.161	0.126	0.202
		SE	0.577	0.292	0.198	0.153	0.126
	β_4	Coeff. Est	1.005	1.029	1.019	1.005	1.013
		Standard Deviation	0.555	0.303	0.199	0.153	0.130
		RMSE	0.555	0.304	0.200	0.153	0.131
		SE	0.574	0.292	0.198	0.152	0.127
PML-FE	β_3	Coeff. Est	0.980	0.947	0.981	0.985	0.977
		Standard Deviation	0.509	0.293	0.183	0.148	0.131
		RMSE	0.508	0.288	0.182	0.147	0.129
		SE	0.501	0.255	0.174	0.137	0.112
	β_4	Coeff. Est	0.949	1.001	0.976	0.966	0.978
		Standard Deviation	0.518	0.324	0.184	0.136	0.134
		RMSE	0.515	0.324	0.182	0.132	0.133
		SE	0.497	0.256	0.174	0.135	0.112

Turning attention to the fixed effects estimators, each performs roughly as expected, providing nearly unbiased estimates of both nearly time-invariant regressors at all levels. At

²⁵Results not reported in text, but available upon request.

the lowest level of within-to-between variation reported, each estimator is quite inefficient. This lends some support to those who have suggested that estimating fixed-effects models on nearly time-invariant risks erroneously rejecting a true relationship (i.e., Type-II error). However, both improve considerably with only minimal increases in the within-to-between variance. Conditional fixed-effects continues to dominate PML-FE with respect to bias, however, the efficiency gains from maintaining the full sample are considerable enough that it is consistently preferred in mean-square error term.

Table 6.2: Panel Estimators with Rarely Changing Regressors ($\alpha_i \sim \text{Bimodal}$)

$\phi=0.25$		$\sigma_{within}/\sigma_{between}$:	0.2	0.4	0.6	0.8	1.0
Pooled	β_3	Coeff. Est	0.764	0.751	0.731	0.719	0.692
		Standard Deviation	0.166	0.148	0.129	0.111	0.095
		RMSE	0.289	0.290	0.298	0.302	0.322
		SE	0.084	0.079	0.073	0.068	0.062
	β_4	Coeff. Est	1.368	1.31	1.23	1.145	1.065
		Standard Deviation	0.160	0.144	0.131	0.110	0.095
		RMSE	0.402	0.342	0.265	0.182	0.115
		SE	0.088	0.084	0.079	0.073	0.067
Random Effects	β_3	Coeff. Est	0.889	0.911	0.947	0.976	0.985
		Standard Deviation	0.216	0.172	0.142	0.119	0.108
		RMSE	0.243	0.194	0.152	0.122	0.109
		SE	0.215	0.169	0.139	0.119	0.105
	β_4	Coeff. Est	1.881	1.588	1.382	1.258	1.183
		Standard Deviation	0.210	0.165	0.141	0.124	0.110
		RMSE	0.905	0.611	0.407	0.286	0.213
		SE	0.212	0.166	0.139	0.120	0.106
Con-FE	β_3	Coeff. Est	1.023	0.998	1.012	1.015	1.014
		Standard Deviation	0.457	0.223	0.165	0.134	0.114
		RMSE	0.457	0.223	0.165	0.134	0.115
		SE	0.462	0.235	0.164	0.131	0.112
	β_4	Coeff. Est	1.033	1.002	1.006	1.011	1.014
		Standard Deviation	0.456	0.232	0.159	0.136	0.114
		RMSE	0.457	0.232	0.159	0.136	0.115
		SE	0.462	0.235	0.164	0.131	0.112
PML-FE	β_3	Coeff. Est	0.989	0.989	0.976	0.995	0.967
		Standard Deviation	0.422	0.422	0.224	0.170	0.111
		RMSE	0.422	0.422	0.223	0.170	0.106
		SE	0.592	0.425	0.218	0.151	0.120
	β_4	Coeff. Est	1.002	1.002	0.954	0.974	0.976
		Standard Deviation	0.486	0.486	0.220	0.150	0.137
		RMSE	0.486	0.486	0.215	0.148	0.135
		SE	0.595	0.426	0.218	0.151	0.120

The same general patterns hold in [Table 6.2](#) when we alter the distribution of the unit effect (e.g., bimodal). The fixed effects estimators perform comparably, if not better than, they did under the ‘normal’ set of experiments. However, the pooled and random effects estimators perform much worse, more severely underestimating the exogenous regressor and overestimating the endogenous one. In particular, the random effects estimator is *significantly* worse when its distributional assumptions are violated. For the endogenous

regressor and low within-to-between variation (the leftmost column) its estimate is nearly 2 times(!) greater than the true value. As such, while slowly-changing regressors seem to indeed complicate analysis, these issues are not unique to fixed effects specifications alone. Indeed, it seems that in many cases failing to properly estimate the unit heterogeneity will often have worse even consequences for our ability to draw accurate inferences about nearly time-invariant regressors.

To consider these risks more systematically, I undertake additional analysis to explore the likelihood of making inferential errors when employing these models. In effect, we are interested in two questions: *i*) what is the risk of a false negative when fixed effects models are estimated? *ii*) what is the risk of a false positive when unit heterogeneity is neglected or mishandled? The former relates to [Beck \(2001\)](#)'s concern that when a variable "changes over time, but slowly, the fixed effects will make it hard for such variables to appear either substantively or statistically significant," while the later is concerned with the possibility of omitted variable bias. To assess the risk of false negatives, I calculate z-statistics from the results presented in [Table 6.1](#) and [Table 6.2](#) and sum the total number of times we would fail to reject the null at the 90% confidence level. For false positives, I run additional experiments where the endogenous slowly-changing variable is omitted from the true data-generating process. That is, X_4 is *only* related to y through their joint correlation to the fixed effect, with any relationship found between the two representing an entirely spurious association. The results for the frequency with which pooled, random effects, and conditional fixed effects produce false negatives and false positives over 1,000 simulations is presented in [Table 6.3](#).

The findings do indicate that fixed-effects estimators have a relatively high risk of false negatives when the within-to-between ratio is very low. Specifically, when within variation is 5 times greater than between variation the null will be incorrectly rejected by fixed-effects estimators 45.5% of the time when the effects are distributed normally and 26.8% of the time when they are distributed bimodally. However, the power of these estimators quickly improves as false-negative rates drop to negligible levels with slight increases in the between-to-within variation, occurring less than 3% of the time when the within variation is 2.5

times greater than the between variation, indicative of a very slow-changing variable. As such, it appears that false negatives, while a concern, occur regularly only when the ratio of between-to-within variation is extremely high. Given that this is an statistical question researchers *can* answer with their data, it seems advisable to calculate the between-to-within variation for all time-varying variables before undertaking any analysis. That is, rather than just assuming particular variables are generally ‘slow moving,’ and forgoing fixed effects, one should actually calculate these descriptive statistics to help inform their model selection as some ‘slow moving’ variables are not negatively impacted by such specifications.

Table 6.3: Type 1 & Type 2 Errors with Slowly-Changing Regressors

$\phi = 0.25$		$\sigma_{within}/\sigma_{between}$:	0.2	0.4	0.6	0.8	1.0
Pooled	<i>Normal</i>	False Negatives	0	0	0	0	0
		False Positives	623	604	589	537	476
	<i>Bimodal</i>	False Negatives	0	0	0	0	0
		False Positives	1000	1000	1000	1000	1000
Random Effects	<i>Normal</i>	False Negatives	0	0	0	0	0
		False Positives	446	387	330	292	221
	<i>Bimodal</i>	False Negatives	0	0	0	0	0
		False Positives	999	994	928	765	627
Con - FE	<i>Normal</i>	False Negatives	455	26	0	0	0
		False Positives	58	50	44	53	43
	<i>Bimodal</i>	False Negatives	268	2	0	0	0
		False Positives	48	58	61	32	60

Simulations run 1000 times. All findings for β_4 using 90 percent confidence level. False negatives are calculated when X_4 is included in the DGP and false positives when it is not.

The pooled and random-effects models always find a relationship where one is present – no false negatives – however, they also regularly conclude support for a relationship where none is present. The pooled estimator erroneously concludes significance nearly 50% of the time for all of the normal experiments, and a full 100% (!!!) of the time in the bimodal experiments. That is, even with a relatively low level of endogeneity specified ($\phi = 0.25$) the pooled estimator consistently – and under some conditions, always – finds significance when we know there to be no direct relationship between the variables. Random effects only offers slight improvements over this, evidencing high false positive rates for all of the specifications

examined. Moreover, the rate at which these estimators incorrectly conclude significance increases with lower within-to-between variation. That is, under the *very* conditions when they are frequently advocated as the preferred alternative.

Whether these type-I or type-II errors are ‘worse’ is not a question I take up here and, moreover, to do so would largely miss the point. Both threaten our ability to make sound inferences and this analysis suggests that the risk of either is significant when our variables are slowly-moving. While, as researchers have noted, there is a threat of false negatives for these regressors when one estimates fixed effects models, the analysis given here has suggested two important addendums. First, the risk of false negatives is only meaningful at very high levels of between-to-within variation – that is, for very slow moving variables – as such researchers should analyze how ‘sluggish’ there variable is rather than treating this as a binary condition. Second, slow-moving variables *also* pose problems for pooled and random effects estimators. As always when unit effects are present, both estimators consistently overestimate the strength of endogenous regressors (and pooled underestimates the strength of exogenous ones), yet these biases are even more severe for slowly-changing regressors. As a consequence the threat of false positives is greatest under the same conditions that researchers fear fixed effects may induce a false negative. While researchers will vary in their preferences for which risk is more acceptable (e.g., how conservative an estimation strategy to adopt), they should not eschew fixed effects under the belief that the estimates obtained by pooled or random effects are sound. To the contrary, these estimators will frequently cause us to erroneously conclude support for relationships where none exist. As such, whichever strategy one ultimately prefers, analysts should be measured in their interpretation of these variables, taking care to note the possibility that their results will be biased – in one direction or the other – for the reasons presented here.

6.3 DEVELOPMENTAL PEACE

In light of these findings, I reanalyze [Fearon & Laitin \(2003\)](#) classic and oft-cited insurgency model which examines the determinants of civil war onset globally from 1945 to 1999. *Civil War* is coded as ‘1’ if there is a civil war – fighting between agents of the state and non-state actors seeking government control, regional autonomy, or changes in government policy in which 1000 deaths occurred over the entire spell with an average of 100 per year and at least 100 on both sides – and ‘0’ otherwise. The covariates are identical to the main model in [Fearon & Laitin \(2003\)](#) and therefore I refer readers there for a more thorough discussion, but these include time-varying regressors (*Population(ln)*, *GDP growth*, *Polity*, *Oil*, *Instability*, *Non-Contig*), several of which are also rarely changing, and time-invariant measures (*logMtn*, *Ethnic Fractionalization*, *Religious Fractionalization*). In addition *War Lag* is included to capture possible temporal dependence.²⁶ Finally, as in [Fearon & Laitin \(2003\)](#), I include GDP per capita (constant in 1985 \$USD), which is logged, *logGDP*. Above I argued that GDP and civil war are jointly determined by an unobserved and unmodeled long-standing dispute resolution capacity, which varies across states. As such, any direct relationship between the two is spurious. To account for this unobserved heterogeneity, I estimate several fixed effects specifications to eliminate these time constant unobservables. If my argument is correct, *GDP(ln)* should no longer be significantly related to conflict once these unobservables are introduced into our model.

In [Table 6.4](#), I estimate models of civil war using pooled logit, random effects, and conditional fixed effects. The results are largely in line with my expectations. As in [Fearon & Laitin \(2003\)](#) the pooled results suggest a negative and significant effect of GDP on civil war. Furthermore, population, mountainous terrain, non-contiguity, oil, and instability are all found to be positively and significantly related to civil war. Concerned with the possibility of unit effects, but wary of the ‘severity’ of fixed effects, I then specify a random effects model. The results are identical to those given in the pooled model, suggesting this

²⁶Again, this is just in keeping with [Fearon & Laitin \(2003\)](#)’s specification, see [chapter 2](#) for a discussion on other approaches to capture temporal dependence in binary data.

specification offers no additional leverage on this question.²⁷ Finally, I estimate a conditional fixed effects model of civil war onset. These results differ markedly from the pooled and random effects estimates, most notably $GDP(ln)$. In model 3, we see that the relationship between GDP and civil war is no longer significant, indicating that there is no significant within unit effect of development on civil war. These results support my expectations and the formal analysis offered by Chassang & Padro-i Miquel (2009), suggesting that the *direct* and casual relationship attributed to development and civil war in the existing literature has been overstated.

Table 6.4: Unit Effects and Civil War

	Model 1	Model 2	Model 3
	Pooled	Random FX	Con - FE
$GDP(ln)$	-0.342*** (0.071)	-0.342*** (0.071)	-0.124 (0.150)
$Pop(ln)$	0.262*** (0.072)	0.262*** (0.072)	1.009** (0.411)
$Mtn\ Est(ln)$	0.219*** (0.085)	0.219*** (0.085)	-
$Ncontig$	0.467* (0.275)	0.467* (0.275)	-
Oil	0.861*** (0.279)	0.861*** (0.279)	0.809 (0.703)
$New\ State$	1.762*** (0.341)	1.762*** (0.341)	1.645*** (0.413)
$Instability$	0.636*** (0.236)	0.636*** (0.236)	0.760*** (0.267)
$Polity2$	0.020 (0.017)	0.020 (0.017)	0.0159 (0.025)
$Eth\ Frac$	0.124 (0.374)	0.124 (0.374)	-
$Rel\ Frac$	0.340 (0.510)	0.340 (0.510)	-
$War\ Lag$	-0.959*** (0.314)	-0.959*** (0.314)	-2.319*** (0.392)
$Constant$	-6.700*** (0.734)	-6.740*** (0.734)	-
$N\ (states)$	6310(143)	6310(143)	2756(56)

SEs in parentheses. ***sig at 1%; **sig at 5%; *sig at 10%.

²⁷This is also confirmed by the non-significant Likelihood Ratio test, $p = 0.497$

However, given the prominence of this finding, it is important to explore the other factors which might explain this non-finding before discarding development as a determinant of conflict. Therefore, I consider the two possible problems regularly associated with fixed effects (as discussed in the previous section). First, scholars have suggested fixed effects is inappropriate in models of conflict because of attendant sample censoring due to the rarity of these events. The argument suggests that our results will be at best inefficient, and at worse biased, by failing to account for those units which ever experience conflict. For GDP in particular, omitting all the units which are high on GDP and low (i.e., zero) on conflict may bias us away from finding a relationship between the two, as we are left to estimate on a sub-set of countries that are all (relatively) impoverished.

I test whether this is indeed the case in [Table 6.5](#). In Model 5, I estimate a pooled logit on the sub-sample of states which have experience civil war at any point during the observational period. If researchers are correct that fixed effects results are biased from “treating as non-informative all countries where we do not observed variation in the response,” this should also present itself in this analysis. However, we see that while the coefficient estimate is attenuated (and the standard error marginally increased), this sample selection does not induce any changes to the sign or significance of GDP. While other variables are affected by this choice – mountain est and non-contig – the estimate of GDP remains positive and significant. Given that this is estimated on the same sample used in fixed effects models, it simply cannot be the case that sample selection alone is inducing the findings we observe. Approaching this same issue from a different direction, in Model 6 I estimate a penalized maximum likelihood fixed effects (PML-FE) model which accounts for individual unit heterogeneity but allows researchers to maintain the entire sample (see [chapter 5](#) for an explanation of this estimator). Here we see that despite maintaining the entire sample in our analysis, the effect of GDP on civil war is no longer significant once unobserved unit effects are accounted for. In sum, the differences in the results given by pooled and fixed effects estimators are not dependent on the sample under analysis, here I have modified the sample for both (censoring pooled and expanding fixed effects) and produced familiar results.

Table 6.5: Unit Effects and Civil War: Sample Size

	Model 4 Full Sample	Model 5 Conflict Only	Model 6 PML - FE
<i>GDP(ln)</i>	-0.342*** (0.071)	-0.172*** (0.076)	-0.715 (0.061)
<i>Pop(ln)</i>	0.262*** (0.072)	0.195*** (0.073)	0.888*** (0.302)
<i>Mtn Est(ln)</i>	0.219*** (0.085)	0.039 (0.091)	-
<i>Ncontig</i>	0.467* (0.275)	0.259 (0.277)	-
<i>Oil</i>	0.861*** (0.279)	0.719*** (0.298)	0.732 (0.521)
<i>New State</i>	1.762*** (0.341)	1.778*** (0.355)	1.552*** (0.335)
<i>Instability</i>	0.636*** (0.236)	0.544*** (0.238)	0.749*** (0.216)
<i>Polity2</i>	0.020 (0.017)	0.023 (0.017)	0.0172 (0.019)
<i>Eth Frac</i>	0.124 (0.374)	-0.173 (0.376)	-
<i>Rel Frac</i>	0.340 (0.510)	0.453 (0.507)	-
<i>War Lag</i>	-0.959*** (0.314)	-1.408*** (0.305)	-2.248*** (0.344)
<i>Constant</i>	-6.700*** (0.734)	-4.998*** (0.734)	-10.100*** (2.402)
<i>N (states)</i>	6310(143)	2772(66)	6310(143)

SEs in parentheses. ***sig at 1%; **sig at 5%; *sig at 10%.

The other possibility presented in the literature is that the slow-changing nature of GDP itself might pose problems for fixed effects estimation. As discussed at length in the last section, when variables change slowly over time removing all the between variation, as in fixed effects, may ‘mask’ the true effect of the regressor. GDP is regularly argued to possess these features, yet no where have I actually seen this analyzed, that is, how slowly does GDP change? This can be easily calculated by computing the mean and standard deviation for GDP within each unit. We can then divide the mean of the standard deviation of GDP within units (0.385) over the standard deviation of the means of GDP across units (0.963)

to produce the within-to-between ratio (0.396). This is actually quite similar to one of the experimental conditions tested earlier $\sigma_{within}/\sigma_{between} = 0.4$, there we saw that under the least favorable conditions reported, fixed effects produces false negatives only 2.6% of the time.²⁸ As such, it is highly unlikely that the fixed effects specification is producing incorrect conclusions.

In sum, the evidence seems to suggest that the fixed effects estimates differ from pooled estimates not because the dependent variable is binary and rare or because GDP changes slowly over time, but because there are unobservables related to both development and civil war. Once these are properly accounted for, using fixed effects, the relationship between GDP and civil war is no longer significant.

6.4 DISCUSSION

Per capita income is widely considered to be the most robust and significant determinant of civil war. In large part, this stems from the undeniable reality that civil wars occur more frequently in low-income countries. However, a strong empirical association does not indicate a casual relationship. Yet, frequently civil war scholars have seemed to conflate the two in the case of development and civil war. Here I have suggested that we need to be cautious in our both our empirical approach and in our claims about this relationship. Specifically, I have advocated the use of fixed effects to study this relationship. Though the language of this chapter has been necessarily methodological, this issue is not simply one of taste or a methodological nuisance. Rather, it represents a significant theoretical belief, namely, whether we believe there are latent unobserved factors which cause some states to experience (avoid) civil conflict. More specifically, are there unobservables which may determine both conflict propensity and the included regressors? I have argued here that there are strong reasons to suspect this is the case from development and civil war.

²⁸We could even take this further and calculate the probability of observing the z-statistic produced in our empirical results here (rather than using traditional levels of significance), doing so would just further support the argument that these findings are unlikely to have arisen from error.

When fixed effects models are estimated development and civil war are no longer significantly related. I have gone to some lengths to show that this *is* a consequence of theoretical differences underlying the models, namely, the presence of unobserved unit heterogeneity. Similar findings have traditionally been dismissed or discounted with appeals made to the shortcomings of fixed effects for such models. While these concerns should be strongly considered and are likely to corrupt *some* relationships, I have shown that they are unlikely to be doing so here. Specifically, neither complications arising from slowly changing regressors nor sample selection are likely to have produced the findings discussed here. Instead, the findings support my argument that unobservable factors contribute to both development and civil war. Once these factors are directly introduced to the systematic part of the model the ‘most robust’ finding in the civil war literature appears to be no more.

In [chapter 3](#) we saw that when the dynamics of civil war are properly accounted for GDP is also no longer significant. As such, we have three possible explanations for what I have called the ‘developmental peace.’ First, that GDP per capita is genuinely a contributing factor to the emergence of civil war, as has been argued consistently in the conflict literature. Second, that conflicts cluster in time and space in a pattern similar to low development, once this temporal and spatial dependence is accounted for there is no longer a direct relationship of GDP on conflict. Finally, that unobservable factors contribute both to the emergence of low GDP and frequent conflict, with no direct relationship between the two. While I have provided evidence for the latter two theories here, we ultimately want to discriminate between the two. As such, I discuss how this might be achieved in the future in [chapter 7](#).

7.0 SUMMARY AND CONCLUSION

While the study of binary rare-events has long been central to the field of International Relations, we have only begun to fully appreciate the unique challenges confronted with these data. Though the problems are often analogous to those confronted with continuous-outcome data, the solutions rarely are. Given the prevalence of these data within a range of important issue areas in political science, it is important to improve our ability to surmount these econometric challenges. Therefore, in the preceding text, I summarized and offered possible solutions for two of these more significant challenges: (i) spatiotemporal auto-dependence and (ii) unit-level heterogeneity. In short, I have engaged issues surrounding the presumed independence of our data. While the problems posed from violating this core tenet of statistical analysis is widely known – see any introductory econometrics book — most applied works remain largely silent on the issues as I have raised them here, with few noting these possible complications and fewer still seeking to redress them.

Spatial, temporal, and spatiotemporal dependence is likely to exist in most any data. While the presence of possible temporal dependence has been appreciated for some time, recent history should *also* motivate us to focus more closely on the means by which crises propagate globally (i.e., spatially). There can be little question as to whether such events are independently determined – they are not – these events are a consequence of both country-level and systemic factors. Failing to incorporate the latter into our analysis risks our ability to draw accurate conclusions, period. That is, failing to account for the (inter-)dependence of these outcomes risks inefficiency at best, and often bias as well. Given these consequences, it will always be good practice to test whether spatial dependence is present, even in those (rare) instances where we expect it to be absent. These concerns are not simply academic, but

have real policy implications for the understanding the occurrence of, and optimal response to, a host of significant political phenomena (e.g., financial crises, civil conflict).

In [chapter 2](#) of my dissertation, I attempted to further our ability to understand such dynamics. I presented a BANC estimator (MSL-by-RIS) for dealing with endogenous spatial lags for binary outcomes. Furthermore, I outline two ways (cumulative multivariate normal probabilities and brute-force simulation) by which we can calculate substantive counterfactual effects, thereby capturing the effect of the occurrence of some event in one unit-time on the probability of an event in another unit-time. Both parameter and, more importantly, effects estimation represent sizable advances over the extant IR ‘spatial’ literature which, with rare exception, fails to address the endogeneity of the spatial lag – treating it as exogenous – and offers no estimate of the substantive effects, which are the quantities we are actually interested in. As such, it is difficult to know what one should conclude from these existing studies, what inference does one draw from a significant coefficient estimate on an endogeneous regressor in a non-linear model? In future research, we should take greater care to ensure our empirical strategies are able to offer meaningful results.

Given that for most of the issues IR scholars are interested in we will almost never have experimental control, we must instead focus on making the most of the the historical data we are presented. This is an incredibly difficult task, rife with complication. Even in the proposed solutions suggested here, I have limited my focus to a particular aspect of interconnectedness; simply looking at whether the level of exposure to the outcomes of others increases the probability of a similar occurrence. In the future this can, and should, be expanded to incorporate (at least) the three legs of the Epidemiologic Triad for disease transmission: *(i)* the virulence of an external agent, *(ii)* the susceptibility of host, *(iii)* and the conditions of the environment. This will give us a much clearer picture on the risk of spatial spillovers than simply looking at the degree of exposure. In future work I plan to extend to attempt to incorporate this theoretic model more completely in our empirical analysis.

Even with the less developed model of interconnectedness presented here, we have seen the clear theoretical importance of understanding these dynamics. In [chapter 3](#), I

showed how the conflict trap is actually a regional phenomena, with spatial and temporal dependence augmenting one another. Specifically, I argue that contagion is a mechanism that increases the persistence of conflict, and more persistent conflicts increase the likelihood of contagion. These mutually-reinforcing positive-feedback loops help explain why we observe conflict-prone regions and offer a more complete understanding of the ‘conflict trap.’ This underlines the importance of a regional approach when peacekeepers are enacting conflict management strategies. In the future it will also be important to understand how the possibility of external intervention influences these dynamics.

My research has also shown the importance of spatial dynamics in understanding financial crises. Specifically, in [chapter 4](#) I find that states with common political fundamentals are more likely to experience simultaneous financial crises. While I argue that this is driven by investor updating, more work is needed to prove that this mechanism should be preferred to possible alternatives. At minimum, however, it suggests that political institutions *do* matter in the contagion of financial crises. Therefore, explaining the exact process by which they matter remains a fruitful area for future research. The results also illustrate the importance of estimating models with a better specified spatial structure. We observed that when included in a model with other possible sources of contagion, the effect of trade is drastically reduced. This lies in stark contrast to much of the existing contagion literature in economics which has largely utilized single spatial-lag models and find strong effects of trade.

In the final two chapters, [chapter 5](#) and [chapter 6](#), I addressed another possible source of dependence in our data: constant unit-effects. While the ‘Dirty Pool’ symposium greatly advanced our understand of how best to model unit-level heterogeneity with binary-outcome data in International Relations, the time had come to revisit that debate. Many of the conclusions rendered there are reconsidered and shown to not be as dire as once thought. Above all, I propose a strategy to avoid the sample losses which generated much of the debate in ‘Dirty Pool’: penalized-maximum-likelihood fixed-effects (PML-FE). I show how PML can provide finite estimates for parameters even in instances of perfect separation, allowing us to estimate a fixed-effects model where all units are retained, minimizing the risk of bias and improving efficiency. As such, I argue, should be the preferred approach for modeling

rare-event binary data. Using this estimator, I (re-)visit the relationship between GDP and civil conflict. I show that despite strong claims to the contrary, we should be suspicious of research arguing for a direct relationship between development and war. Once we account for potential unobservables, driving both development and conflict, there is no longer any relationship. While researchers have called upon the ‘Dirty Pool’ conclusions to explain away this non-finding, I show that many of these justifications are specious (e.g., sample selection, nearly time-invariant regressors) and that fixed effects should be the preferred approach. As such, I we need to reconsider the debate linking development and poverty and offer alternative explanations for what remains an interesting empirical regularity.

8.0 BIBLIOGRAPHY

- Acemoglu, Daron, Johnson, Simon, Robinson, James A, & Yared, Pierre. 2008. Income and Democracy. *American Economic Review*, **98**(3), 808–842.
- Achen, Christopher H. 2002. Toward a new political methodology: Microfoundations and ART. *Annual Review of Political Science*, **5**(1), 423–450.
- Agenor, Pierre-Richard. 2003. Benefits and Costs of International Financial Integration: Theory and Facts. *The World Economy*, **26**(8), 1089–1118.
- Albert, Adelin, & Anderson, JA. 1984. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, **71**(1), 1–10.
- Allen, Franklin, & Gale, Douglas. 2000. Financial Contagion. *Journal of Political Economy*, **108**, 1–33.
- Allen, Franklin, & Gale, Douglas. 2007. *Understanding Financial Crises*.
- Allen, Franklin, Babus, Ana, & Carletti, Elena. 2009. Financial Crises: Theory and Evidence. *Annual Review of Financial Economics*, **1**, 97–116.
- Angrist, Joshua D, & Pischke, Jörn-Steffen. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Athukorala, Prema-chandra, & Warr, Peter. 2002. Vulnerability to a Currency Crisis: Lesson from the Asian Experience. *World Economy*, **25**(1), 33–57.

- Barber, Brad M, & Odean, Terrance. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, **21**(2), 785–818.
- Barbieri, Katherine, Keshk, Omar MG, & Pollins, Brian M. 2009. Trading data evaluating our assumptions and coding rules. *Conflict Management and Peace Science*, **26**(5), 471–491.
- Barbieri, Katherine, Keshk, Omar MG, & Pollins, Brian M. 2012. Correlates of War Project Trade Data Set Codebook, Version 3.0. <http://correlatesofwar.org>.
- Beck, N., Katz, J., & Tucker, R. 1998. Taking time seriously: Time-series cross-section analysis with a binary dependent variable. *American Journal of Political Science*, **42**(4), 1260–88.
- Beck, N., Gleditsch, K., & Beardsley, K. 2006. Space is More than Geography: Using Spatial Econometrics in the Study of Political Economy. *International Studies Quarterly*, **50**(1), 27–44.
- Beck, Nathaniel. 2001. Time-series-cross-section data: what have we learned in the past few years? *Annual review of political science*, **4**(1), 271–293.
- Beck, Nathaniel. 2008. Time-series-cross-section methods. *Oxford Handbook of Political Methodology*, 475–93.
- Beck, Nathaniel. 2011. Of fixed-effects and time-invariant variables. *Political Analysis*, **19**(2), 119–122.
- Beck, Nathaniel, & Katz, Jonathan N. 2001. Throwing out the baby with the bath water: A comment on Green, Kim, and Yoon. *International Organization*, **55**(2), 487–495.
- Beck, Nathaniel, & Katz, Jonathan N. 2007. Random coefficient models for time-series–cross-section data: Monte Carlo experiments. *Political Analysis*, **15**(2), 182–195.

- Beck, Nathaniel, Epstein, David L, Jackman, Simon, & O'Halloran, Sharyn L. 2001. Alternative models of dynamics in binary time-series-cross-section models: The example of state failure.
- Bekaert, Geert, Harvey, Campbell R, & Lundblad, Christian. 2005. Does financial liberalization spur growth? *Journal of Financial economics*, **77**(1), 3–55.
- Beron, Kurt J, & Vijverberg, Wim PM. 2004. Probit in a spatial context: a Monte Carlo analysis. *Pages 169–195 of: Advances in spatial econometrics*. Springer.
- Beron, Kurt J, Murdoch, James C, & Vijverberg, Wim PM. 2003. Why cooperate? Public goods, economic power, and the Montreal Protocol. *Review of Economics and Statistics*, **85**(2), 286–297.
- Bester, C Alan, & Hansen, Christian B. 2013. Grouped effects estimators in fixed effects models. *Journal of Econometrics*.
- Blattman, Christopher, & Miguel, Edward. 2010. Civil war. *Journal of Economic Literature*, 3–57.
- Bonhomme, Stéphane, & Manresa, Elena. 2012. Grouped Patterns of Heterogeneity in Panel Data.
- Bordo, Michael, Eichengreen, Bordo, Klingebiel, Daniela, & Martinez-Peria, Maria Soledad. 2001. Is the Crisis Problem Growing More Severe? *Economic Policy*, **16**(32), 51–82.
- Boyd, John H, Kwak, Sungkyu, & Smith, Bruce. 2005. The real output losses associated with modern banking crises. *Journal of Money, Credit and Banking*, 977–999.
- Buhaug, Halvard, & Gleditsch, Kristian Skrede. 2008. Contagion or Confusion? Why Conflicts Cluster in Space1. *International Studies Quarterly*, **52**(2), 215–233.
- Büthe, Tim, & Milner, Helen V. 2008. The politics of foreign direct investment into developing countries: increasing FDI through international trade agreements? *American Journal of Political Science*, **52**(4), 741–762.

- Calvo, Guillermo, & Mendoza, Enrique G. 2000. Rational Contagion and the Globalization of Securities Markets. *Journal of International Economics*, **51**(1), 79–113.
- Calvo, Sara, & Reinhart, Carmen M. 1996. Capital Flows to Latin America: Is There Evidence of Contagion Effects? *In: Private Capital Flows to Emerging Markets Washington DC: Institute for International Economics.*
- Carter, D. B., & Signorino, C. S. 2010. Back to the Future: Modeling Time Dependence in Binary Data. *Political Analysis*, **18**(3), 271–292.
- Chassang, Sylvain, & Padro-i Miquel, Gerard. 2009. Economic shocks and civil war. *Quarterly Journal of Political Science*, **4**(3), 211–228.
- Cheibub, José Antonio, Gandhi, Jennifer, & Vreeland, James Raymond. 2010. Democracy and dictatorship revisited. *Public Choice*, **143**(1-2), 67–101.
- Chossudovsky, Michel. 1996. Economic genocide in Rwanda. *Economic and Political Weekly*, 938–941.
- Collier, P., & Hoeffler, A. 2004. Greed and Grievance in Civil War. *Oxford Economic Papers*, **56**, 563–595.
- Collier, P., Elliot, L., Hegre, H., Hoeffler, A., Reynal-Querol, M., & Sambanis, N. 2003. *Breaking the Conflict Trap. Civil War and Development Policy.* Oxford, UK: Oxford University Press.
- Collier, Paul. 2000. Rebellion as a quasi-criminal activity. *Journal of Conflict Resolution*, **44**(6), 839–853.
- Collier, Paul. 2008. *The Bottom Billion: Why the Poorest Countries Are Failing and What Can Be Done About It.* New York, NY: Oxford University Press.
- Collier, Paul, Hoeffler, Anke, & Söderbom, Måns. 2008. Post-conflict risks. *Journal of Peace Research*, **45**(4), 461–478.

- Collier, Paul, Hoeffler, Anke, & Rohner, Dominic. 2009. Beyond greed and grievance: feasibility and civil war. *Oxford Economic Papers*, **61**(1), 1–27.
- Cook, Scott, Hays, Jude, & Franzese, Robert. 2014. *Estimating Unit Effects with Binary Data: A Penalized Maximum Likelihood Approach. Working Paper.*
- Cordeiro, Gauss M, & McCullagh, Peter. 1991. Bias correction in generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 629–643.
- Corsetti, Ginacarlo, Pesenti, Paolo, Roubini, Nouriel, & Tille, Cedric. 2000. Competitive devaluations: toward a welfare-based approach. *Journal of International Economics*, **51**(1), 217–214.
- Cramer, JS. 2005. Omitted Variables and Misspecified Disturbances in the Logit Model. *Rand Journal of Economics*, **44**(3), 391–424.
- Dafoe, Allan, Oneal, John R, & Russett, Bruce. 2013. The democratic peace: Weighing the evidence and cautious inference. *International Studies Quarterly*, **57**(1), 201–214.
- Daude, Christian, & Stein, Ernesto. 2007. The Quality of Institutions and Foreign Direct Investment. *Economics & Politics*, **19**(3), 317–344.
- De Gregorio, Jose, & Valdes, Rodrigo. 2001. Crisis Transmission: Evidence from the Debt, Tequila, and Asian Flu Crises. *World Bank Economic Review*, **15**(2), 289–314.
- De Mesquita, Bruce Bueno, Morrow, James D, Siverson, Randolph M, & Smith, Alastair. 1999. An institutional explanation of the democratic peace. *American Political Science Review*, 791–807.
- Diehl, P. F. 1991. Geography and war: a review and assessment of the empirical literature. *International Interactions*, **17**, 11–27.
- Djankov, Simeon, & Reynal-Querol, Marta. 2010. Poverty and civil war: Revisiting the evidence. *The Review of Economics and Statistics*, **92**(4), 1035–1041.

- Dornbush, Rudiger, Yung Chul Park, & Claessens, Stijn. 2000. Contagion: Understanding How It Spreads. *The World Bank Research Observer*, **15**(2), 177–197.
- Doyle, Michael W. 1986. Liberalism and world politics. *American Political Science Review*, **80**(4), 1151–1169.
- Dungey, Mardi, & Tambakis, Demosthenes. 2005. Speculative Attacks on Pegged Exchange Rates. *Pages 3–33 of: Dungey, Mardi, & Tambakis, Demosthenes (eds), Identifying International Financial Contagion: Progress and Challenges*. New York, NY: Oxford University Press.
- Dungey, Mardi, Fry, Renee, & Martin, Vance. 2006. Correlations, Contagion, and Asian Evidence. *Asian Economic Papers*, **5**(2), 33–67.
- Elbadawi, Ibrahim, & Sambanis, Nicholas. 2002. How much war will we see? Explaining the prevalence of civil war. *Journal of Conflict Resolution*, **46**(3), 307–334.
- Fearon, James. 2008. Economic development, insurgency, and civil war. *In: Institutions and economic performance*. Cambridge, MA: Harvard University Press.
- Fearon, James D. 1995. Rationalist explanations for war. *International Organization*, **49**, 379–379.
- Fearon, James D., & Laitin, David D. 2003. Ethnicity, Insurgency, and Civil War. *American Political Science Review*, **97**(1), 75–90.
- Feng, Yi. 2001. Political Freedom, Political Instability, and Policy Uncertainty: A Study of Political Institutions and Private Investment in Developing Countries. *International Studies Quarterly*, **45**(2), 271–294.
- Firth, David. 1993. Bias reduction of maximum likelihood estimates. *Biometrika*, **80**(1), 27–38.

- Fjelde, Hanne, & De Soysa, Indra. 2009. Coercion, Co-optation, or Cooperation? State Capacity and the Risk of Civil War, 1961-2004*. *Conflict Management and Peace Science*, **26**(1), 5–25.
- Folkes, Valerie S. 1988. Recent attribution research in consumer behavior: A review and new directions. *Journal of Consumer Research*, 548–565.
- Frankel, Jeffrey A. 2005. *Contractionary currency crashes in developing countries*. Tech. rept. National Bureau of Economic Research.
- Franzese, Robert, Hays, Jude, & Cook, Scott. 2014. *Spatial- and Spatiotemporal-Autoregressive Probit Models of Interdependent Binary Outcomes*. Working Paper.
- Franzese, Robert J, & Hays, Jude C. 2007. Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data. *Political Analysis*, **15**(2), 140–164.
- Franzese, Robert J, & Hays, Jude C. 2008. Spatial Analysis. *Oxford Handbook of Political Methodology*, 570–604.
- Furfine, Craig. 2003. Interbank Exposures: Quantifying the Risk of Contagion. *Journal of Money, Credit and Banking*, **35**(1), 118–128.
- Gartzke, Erik. 2007. The capitalist peace. *American Journal of Political Science*, **51**(1), 166–191.
- Gates, Scott. 2002. Empirically assessing the causes of civil war. *Pages 24–27 of: Convention of the 43rd Annual International Studies Association, New Orleans, LA*.
- Gleditsch, Kristian Skrede. 2002. *All international politics is local: The diffusion of conflict, integration, and democratization*. University of Michigan Press.
- Glick, Reuven, & Rose, Andrew K. 1999. Contagion and Trade: Why are Currency Crises Regional. *Journal of International Money & Finance*, **18**(4), 603–618.

- Goldstein, Morris. 1998. *The Asian financial crisis: Causes, cures, and systemic implications*. Vol. 55. Peterson Institute.
- Gorton, Gary. 1988. Banking Panics and Business Cycles. *Oxford Economic Papers*, **40**(4), 751–781.
- Gowa, Joanne. 1994. *Allies, adversaries, and international trade*. Princeton University Press.
- Gowa, Joanne. 1995. Democratic states and international disputes. *International Organization*, 511–522.
- Green, Donald P, Kim, Soo Yeon, & Yoon, David H. 2001. Dirty pool. *International Organization*, **55**(2), 441–468.
- Greene, William. 2004. The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal*, **7**(1), 98–119.
- Grossman, Herschel I. 1991. A general equilibrium model of insurrections. *American Economic Review*, **81**(4), 912–21.
- Hafner-Burton, Emilie M. 2005. Trading human rights: How preferential trade agreements influence government repression. *International Organization*, 593–629.
- Haggard, Stephan. 2000. *The Political Economy of the Asian Financial Crisis*. Washington: Institute for International Economics.
- Haggard, Stephan, & McCubbins, Matthew D. 2001. *Presidents, parliaments, and policy*. Cambridge University Press.
- Harbom, Lotta, & Wallensteen, Peter. 2007. Armed Conflict, 1989–2006. *Journal of Peace Research*, **44**(5), 623–634.

- Hausman, Jerry A, Abrevaya, Jason, & Scott-Morton, Fiona M. 1998. Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics*, **87**(2), 239–269.
- Hays, J. 2009. Bucking the System: Using Simulation Methods to Estimate and Analyze Systems of Equations with Qualitative and Limited Dependent Variables. In: *2nd Annual St. Louis Area Methods Meeting (SLAMM), Washington University in St. Louis* ([https://netfiles.uiuc.edu/jchays/www/Hays SLAMM09 Revised. pdf](https://netfiles.uiuc.edu/jchays/www/Hays%20SLAMM09%20Revised.pdf)).
- Hays, Jude C., Freeman, John R., & Nesseth, Hans. 2003. Exchange Rate Volatility and Democratization in Emerging Market Countries. *International Studies Quarterly*, **47**(2), 203–228.
- Hays, Jude C, Kachi, Aya, & Franzese, Robert J. 2010. A spatial model incorporating dynamic, endogenous network interdependence: A political science application. *Statistical Methodology*, **7**(3), 406–428.
- Heckman, J. J. 1978. Dummy Endogenous Variables in a Simultaneous Equation System. *Econometrica*, **46**, 931–59.
- Heckman, James J. 1981. Heterogeneity and state dependence. *Pages 91–140 of: Studies in labor markets*. University of Chicago Press.
- Hegre, H., & Sambanis, N. 2006. Sensitivity Analysis of Empirical Results of Civil War Onset. *Journal of Conflict Resolution*, **50**(4), 508–535.
- Hegre, H., Nygard, H., Strand, H., Gates, S., & Flaten, R. 2011. *The Conflict Trap*. Working Paper. .
- Hegre, Håvard. 2001. Toward a democratic civil peace? Democracy, political change, and civil war, 1816–1992. **95**(01), 33–48.
- Heinze, Georg. 2006. A comparative investigation of methods for logistic regression with separated or nearly separated data. *Statistics in medicine*, **25**(24), 4216–4226.

- Heinze, Georg, & Schemper, Michael. 2002. A solution to the problem of separation in logistic regression. *Statistics in medicine*, **21**(16), 2409–2419.
- Helleiner, Eric. 2011. Understanding the 2007-2008 global financial crisis: Lessons for scholars of international political economy. *Annual Review of Political Science*, **14**, 67–87.
- Henisz, W. J. 2000. The Institutional Environment for Economic Growth. *Economics & Politics*, **12**(1), 1–31.
- Herring, R.J. 1999. Credit risk and financial instability. *Oxford Review of Economic Policy*, **15**(3), 63–79.
- Heston, Alan, Summers, Robert, & Aten, Bettina. 2012 (July). *Penn World Table Version 7.1*. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.
- Hill, Claire A. 1998. How Investors React To Political Risk. *Duke Journal of Comparative & International Law*, **8**(2), 283–313.
- Hirshleifer, Jack. 1995. Anarchy and its breakdown. *Journal of Political Economy*, 26–52.
- Holtermann, Helge. 2012. Explaining the Development–Civil War Relationship. *Conflict Management and Peace Science*, **29**(1), 56–78.
- Hutchinson, Michael, & Noy, Ilan. 2006. Sudden stop and the Mexican wave: Currency crises, capital flow reversals and output loss in emerging markets. *Journal of Development Economics*, **79**(1), 225–248.
- Jackman, S. 2000. *In and Out of War and Peace: Transitional Models of International Conflict*. Unpublished. Stanford University.
- Jensen, Nathan. 2003. Democratic Governance and Multinational Corporations: The Political Economy of Foreign Direct Investment. *International Organizations*, **57**(3), 587–616.

- Jensen, Nathan. 2006. *Nation States and the Multinational Corporation: A Political Economy of Foreign Direct Investment* Princeton. NJ: Princeton University Press.
- Joyce, Joseph, & Nabar, Malhar. 2009. Sudden stops, banking crises and investment collapses in emerging markets. *Journal of Development Economics*, **90**(2), 314–322.
- Justino, Patricia. 2006. On the links between violent conflict and chronic poverty: How much do we really know? *Chronic Poverty Research Centre Working Paper*.
- Kaminsky, Graciela, & Schmukler, S. 1999. *Short and Long-Run Integration: Do Capital Controls Matter?* Working Paper. PRE.
- Kaminsky, Graciela L., & Reinhart, Carmen M. 2000. On Crises, Contagion, and Confusion. *Journal of International Economics*, **51**(1), 145–168.
- Kaminsky, Graciela L, Reinhart, Carmen, & Vegh, Carlos A. 2003. *The unholy trinity of financial contagion*. Tech. rept. National Bureau of Economic Research.
- Karolyi, G. Andrews. 2003. Does International Financial Contagion Exist? *International Finance*, **6**(2), 179–99.
- Kathman, Jacob D. 2010. Civil War Contagion and Neighboring Interventions¹. *International Studies Quarterly*, **54**(4), 989–1012.
- Kathman, Jacob D. 2011. Civil war diffusion and regional motivations for intervention. *Journal of Conflict Resolution*, **55**(6), 847–876.
- Kindleberger, Charles P. 1978. *Manias, Panics, and Crashes: A History of Financial Crisis 1st edition*. New York, NY: Basic Books.
- King, Gary. 2001. Proper nouns and methodological propriety: Pooling dyads in international relations data. *International Organization*, **55**(2), 497–507.
- King, Gary, & Zeng, Langche. 2001a. Explaining rare events in international relations. *International Organization*, **55**(3), 693–715.

- King, Gary, & Zeng, Langche. 2001b. Logistic regression in rare events data. *Political analysis*, **9**(2), 137–163.
- King, MA, & Wadwani, S. 1990. Transmission of Volatility Between Stock Markets. *Review of Financial Studies*, **3**(1), 5–33.
- Kodres, Laura E, & Pritsker, Matthew. 2002. A rational expectations model of financial contagion. *The Journal of Finance*, **57**(2), 769–799.
- Kosmidis, Ioannis. 2007. Bias Reduction in Exponential Family Nonlinear Models.
- Kosmidis, Ioannis, & Firth, David. 2009. Bias reduction in exponential family nonlinear models. *Biometrika*, **96**(4), 793–804.
- Laeven, Luc, & Valencia, Fabian. 2012. *Systemic Banking Crises: A New Database*. IMF Working Paper.
- Lagoarde-Segot, Thomas, & Lucey, Brian M. 2006. Financial vulnerability in emerging markets. Evidence from the Middle East and North Africa. *The Institute for International Integration Studies Discussion Paper Series*, *iisdp114*.
- Lake, David A, & Rothchild, Donald S. 1998. *The international spread of ethnic conflict: Fear, diffusion, and escalation*. Princeton University Press.
- Leblang, David, & Satyanath, Shanker. 2006. Institutions, expectations, and currency crises. *International Organization*, **60**(1), 245.
- Lee, Byunghwan, O'Brien, John, & Sivaramakrishnan, Konduru. 2008. An analysis of financial analysts' optimism in long-term growth forecasts. *The Journal of Behavioral Finance*, **9**(3), 171–184.
- Levy, Jack S. 1988. Domestic politics and war. *Journal of Interdisciplinary History*, 653–673.
- Li, Quan. 2006. Democracy, Autocracy, and Tax Incentives to Foreign Direct Investors: A Cross-National Analysis. *Journal of Politics*, **68**(1), 62–74.

- Li, Quan, & Resnik, Adam. 2003. Reversal of Fortunes: Democracy, Property Rights and Foreign Direct Investment Inflows in Developing Countries. *International Organizations*, **57**(1), 175–214.
- Macias, Jose, Brambila, Isabella Massa, & Salois, Matthew J. 2010. The Impact of Financial Crises on Trade Flows: A Developing Country Perspective. *Presented at the*.
- MacIntyre, Andrew. 2001. Institutions and Investors: The Politics of the Economics Crisis in Southeast Asia. *International Organization*, **55**(1), 81–122.
- Maddala, George S. 1987. Limited dependent variable models using panel data. *Journal of Human Resources*, 307–338.
- Magee, Christopher SP. 2008. New measures of trade creation and trade diversion. *Journal of International Economics*, **75**(2), 349–362.
- Mansfield, Edward, & Milner, Helen. 1999. The new wave of regionalism. *International organization*, **53**(3), 589–627.
- Mansfield, Edward D, & Pevehouse, Jon C. 2000. Trade blocs, trade flows, and international conflict. *International organization*, **54**(4), 775–808.
- Mansfield, Edward D, & Reinhardt, Eric. 2003. Multilateral determinants of regionalism: The effects of GATT/WTO on the formation of preferential trading arrangements. *International organization*, **57**(4), 829–862.
- Mansfield, Edward D, Milner, Helen V, & Rosendorff, B Peter. 2002. Why democracies cooperate more: Electoral control and international trade agreements. *International Organization*, **56**(3), 477–513.
- Mansfield, Edward D, Milner, Helen V, & Pevehouse, Jon C. 2007. Vetoing co-operation: the impact of veto players on preferential trading arrangements. *British Journal of Political Science*, **37**(3), 403.

- Maoz, Zeev, & Russett, Bruce. 1993. Normative and structural causes of democratic peace. *American Political Science Review*, **87**(3), 624–38.
- Masoon, Paul. 2000. Multiple Equilibria, Contagion, and the Emerging Markets. *In: Financial Crises in Emerging Markets eds.* Cambridge: Cambridge University Press.
- McHugh, Kristin. 2005. *Civil Wars: How the World Suffers*.
- Morgenstern, O. 1959. *International Financial Transactions and the Business Cycle* Princeton. NJ: Princeton University Press.
- Moser, Thomas. 2003. What Is International Financial Contagion. *International Finance*, **6**(2), 157–178.
- Mosley, Layna, & Singer, David Andrew. 2009. The global financial crisis: lessons and opportunities for international political economy.
- Most, B., & Starr, H. 1980. Diffusion, Reinforcement, and Geopolitics and the Spread of War. *American Political Science Review*, **74**(4), 932–946.
- Murdoch, J., & Sandler, T. 2002. Economic growth, civil wars, and spatial spillovers. *Journal of Conflict Resolution*, **46**(1), 91–110.
- Nel, Philip, & Righarts, Marjolein. 2008. Natural disasters and the risk of violent civil conflict. *International Studies Quarterly*, **52**(1), 159–185.
- Nickell, Stephen. 1981. Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, 1417–1426.
- Novo, Alvaro. 2003. *Contagious Currency Crises: A Spatial Probit Approach*. Banco de Portugal Working Paper.
- O’Loughlin, J., & Raleigh, C. 2008. Spatial analysis of civil war violence. *In: Cox, R., Low, M. M., & Robinson, J. (eds), The Sage Handbook of Political Geography*. Thousand Oaks, CA: Sage.

- Oneal, John R, & Russett, Bruce. 2001. Clear and clean: The fixed effects of the liberal peace. *International Organization*, **55**(2), 469–485.
- Pasquariello, Paolo. 2007. Imperfect competition, information heterogeneity, and financial contagion. *Review of Financial Studies*, **20**(2), 391–426.
- Pericoli, Marcello, & Sbracia, Massimo. 2003. A Primer on Financial Contagion. *Journal of Economic Surveys*, **17**(4), 571–608.
- Plümper, Thomas, & Troeger, Vera E. 2007. Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis*, **15**(2), 124–139.
- Poirier, Dale J, & Ruud, Paul A. 1988. Probit with dependent observations. *The Review of Economic Studies*, **55**(4), 593–614.
- Powell, Robert. 2006. War as a commitment problem. *International Organization*, **60**(1), 169.
- Pritsker, M. 2001. The Channels for Financial Contagion. *In*: S, Claessens, & Forbes, K. (eds), *International Financial Contagion*. Kluwer Academic Publishers.
- Quinn, Dennis P., & Woolley, John T. 2001. Democracy and National Economic Performance: The Preference for Stability. *American Journal of Political Science*, **45**(3), 634–657.
- Reinhart, Carmen M., & Rogoff, Kenneth. 2011. *This Time IS Different: Eight Centuries of Financial Folly Princeton*. Princeton University Press: NJ.
- Rice, Susan, Graff, Corinne, & Lewis, Janet. 2006. Poverty and civil war: what policymakers need to know. *Brookings Global Economy and Development Working Paper*.
- Rodrik, Dani. 1998. *Where Did All the Growth Go? External Shocks, Social Conflict, and Growth Collapses*. CEPR Discussion Papers.

- Rosato, Sebastian. 2003. The flawed logic of democratic peace theory. *American Political Science Review*, **97**(04), 585–602.
- Russett, Bruce. 1994. *Grasping the democratic peace: Principles for a post-Cold War world*. Princeton University Press.
- Salehyan, I., & Gleditsch, K. S. 2006. Refugees and the spread of civil war. *International Organization*, **60**(2), 335–366.
- Sambanis, Nicholas. 2002. A review of recent advances and future directions in the quantitative literature on civil war. *Defence and Peace Economics*, **13**(3), 215–243.
- Sambanis, Nicholas. 2004. What is Civil War? Conceptual and Empirical Complexities of an Operational Definition. *Journal of Conflict Resolution*, **48**(6), 814–858.
- Santner, Thomas J, & Duffy, Diane E. 1986. A note on A. Albert and JA Anderson’s conditions for the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, **73**(3), 755–758.
- Schaefer, Robert L. 1983. Bias correction in maximum likelihood logistic regression. *Statistics in Medicine*, **2**(1), 71–78.
- Schneider, Christina. 2013. Bilateral Financial Bailouts. *Working paper*.
- Shiller, Robert J. 2005. *Irrational exuberance*. Random House LLC.
- Simmons, Beth A, Dobbin, Frank, & Garrett, Geoffrey. 2006. Introduction: The International Diffusion of Liberalism. *International Organization*, **60**(4).
- Skaperdas, Stergios. 1992. Cooperation, conflict, and power in the absence of property rights. *American Economic Review*, **82**(4), 720–739.
- Spiro, David E. 1994. The insignificance of the liberal peace. *International Security*, **19**(2), 50–86.

- Starr, H., & Most, B. 1983. Contagion and Border Effects on Contemporary African Conflict. *Comparative Political Studies*, **16**(1), 92–117.
- Stiglitz, Joseph E. 2002. Information and the Change in Paradigm in Economics. *The American Economic Review*, **92**(3), 460–501.
- Stinnett, Douglas M, Tir, Jaroslav, Diehl, Paul F, Schafer, Philip, & Gochman, Charles. 2002. The Correlates of War (COW) project direct contiguity data, version 3.0. *Conflict Management and Peace Science*, **19**(2), 59–67.
- Suhrke, A., & Samset, I. 2007. What’s in a Figure? Estimating Recurrence of Civil War. *International Peacekeeping*, **14**(2), 195–203.
- Themnér, Lotta, & Wallensteen, Peter. 2013. Armed Conflicts, 1946–2012. *Journal of Peace Research*, **50**(4), 509–521.
- Thies, Cameron G. 2010. Of rulers, rebels, and revenue: State capacity, civil war onset, and primary commodities. *Journal of Peace Research*, **47**(3), 321–332.
- Train, Kenneth. 2009. *Discrete choice methods with simulation*. Cambridge university press.
- Tsebelis, George. 1995. Decision Making in Political Systems: Veto Players in Presidentialism, Parliamentarism, Multicameralism and Multipartyism. *British Journal of Political Science*, **25**(3), 289–235.
- Tversky, Amos, & Kahnemen, Daniel. 1973. Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, **5**(2), 207–232.
- Upper, Christian. 1996. *Contagion Due to Interbank Credit Exposures: What Do We Know, Why Do We Know It, and What Should We Know?* Bank for International Settlements, working paper.
- Walters, B. 2004. Does Conflict Beget Conflict? Explaining Recurring Civil War. *Journal of Peace Research*, **41**(3), 371–388.

- Ward, M., & Gledistch, K. 2002. Location, location, location: An MCMC approach to modeling the spatial context of war and peace. *Political Analysis*, **10**(3), 244–260.
- Wilhelm, Stefan, & de Matos, Miguel Godinho. 2013. Estimating Spatial Probit Models in R. *R Journal*, **5**(1).
- Wooldridge, Jeffrey. 2012. *Introductory econometrics: A modern approach*. Cengage Learning.
- Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. MIT press.
- Xiang, Jun. 2010. Relevance as a latent variable in dyadic analysis of conflict. *The Journal of Politics*, **72**(02), 484–498.
- Zorn, Christopher. 2005. A solution to separation in binary response models. *Political Analysis*, **13**(2), 157–170.